# PLOS Ten Simple Rules

“Ten Simple Rules” provide a quick, concentrated guide for mastering some of the professional challenges research scientists face in their careers.

# Contents

## Editorials



## Career Development and Choices

- Ten Simple Rules for Building and Maintaining a Scientific Reputation

  June 30, 2011 Philip Bourne, Virginia Barbour

- Ten Simple Rules for Aspiring Scientists in a Low-Income Country

  May 30, 2008 Edgardo Moreno, José-María Gutiérrez

- Ten Simple Rules for Selecting a Postdoctoral Position

  November 24, 2006 Philip E Bourne, Iddo Friedberg

- Ten Simple Rules for Getting Grants

  February 24, 2006 Philip E Bourne, Leo M Chalupa

## Collaboration from Local to International

- Ten simple rules for helping newcomers become contributors to open projects

  September 12, 2019 Dan Sholler, Igor Steinmacher, Denae Ford, Mara Averick, Mike Hoye, Greg Wilson

- Ten simple rules towards healthier research labs

  April 11, 2019 Fernando T. Maestre

- Ten Simple Rules for avoiding and resolving conflicts with your colleagues

  January 18, 2019 Fran Lewitter, Philip E. Bourne, Teresa K. Attwood

- Ten simple rules for international short-term research stays

  December 7, 2017 Diego A. Forero, Sandra Lopez-Leon, George P. Patrinos

- Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing

  January 19, 2017 Mary Regina Boland, Konrad J. Karczewski, Nicholas P. Tatonetti

- Ten Simple Rules for a Successful Cross-Disciplinary Collaboration

  April 30, 2015 Bernhard Knapp, Remi Bardenet, Miguel Bernabeu, Rafel Bordas, Maria Bruna, Ben Calderhead, Jonathan Cooper, Alexander Fletcher, Derek Groen, Bram Kijuper, Joanna Lewis, Greg McInerny, Timo Minssen, James Osborne, Verena Paulitschke, Joe Pitt-Francis, Jelena Todoric, Christian Yates, David Gavaghan, Charlotte Deane

- Ten Simple Rules for Cultivating Open Science and Collaborative R&D

  September 26, 2013 Hassan Masum, Aarthi Rao, Benjamin Good, Matthew Todd, Aled Edwards, Leslie Chan, Barry Bunin, Andrew Su, Zakir Thomas, Philip Bourne

- Ten Simple Rules for Getting Help from Online Scientific Communities

  September 29, 2011 Giovanni Dall'Olio, Jacopo Marino, Michael Schubert, Kevin Keys, Melanie Stefan, Colin Gillespie, Pierre Poulain, Khader Shameer, Robert Sugar, Brandon Invergo, Lars Jensen, Jaume Bertranpetit, Hafid Laayouni

- Ten Simple Rules for a Successful Collaboration

  March 30, 2007 Quentin Vicens, Philip E Bourne

## Education and Mentoring

- Ten simple rules for successfully hosting an intern at a scientific software company

  June 20, 2019 Kristine Briedis, Yi-Shiou Chen, Scott Markel

- Ten simple rules for providing a meaningful research experience to high school students

  April 25, 2019 Emily A. Lescak, Kate M. O'Neill, Giovanna M. Collu, Subhamoy Das

- Ten simple rules for delivering live distance training in bioinformatics across the globe using webinars

  November 15, 2018 Denise Carvalho-Silva, Leyla Garcia, Sarah L. Morgan, Cath Brooksbank, Ian Dunham

- Ten simple rules to create a serious game, illustrated with examples from structural biology

  March 8, 2018 Marc Baaden, Olivier Delalande, Nicolas Ferey, Samuela Pasquali, Jérôme Waldispühl, Antoine Taly

- Ten simple rules for collaborative lesson development

  March 1, 2018 Gabriel A. Devenyi, Rémi Emonet, Rayna M. Harris, Kate L. Hertweck, Damien Irving, Ian Milligan, Greg Wilson

- Ten simple rules for developing a mentor–mentee expectations document

  September 21, 2017 Kristyn S. Masters, Pamela K. Kreeger

- Ten simple rules for successfully completing a graduate degree in Latin America

  September 21, 2017 Ernesto Ruelas Inzunza, Gabriela I. Salazar-Rivera, Magdiel Láinez, María Guadalupe Ruiz-Gómez, Carlo A. Domínguez-Eusebio, Griselda Cristóbal-Sánchez, Issaac A. Teodosio Faustino, Edel Pérez-López, Meagan L. Campbell, Marcus Vinicius Merfa, Ivonne Tatiana Latorre Beltrán, Fernanda Armas, Claudio Mota-Vargas

- Ten simple rules for surviving an interdisciplinary PhD

  May 25, 2017 Samuel Demharter, Nicholas Pearce, Kylie Beattie, Isabel Frost, Jinwoo Leem, Alistair Martin, Robert Oppenheimer, Cristian Regep, Tammo Rukat, Alexander Skates, Nicola Trendel, David Gavaghan, Charlotte Deane, Bernhard Knapp

- Ten Simple Rules for Developing a MOOC

  October 20, 2016 David T. Manallack, Elizabeth Yuriev

- Ten Simple Rules for Finishing Your PhD

  December 4, 2014 Jacopo Marino, Melanie I. Stefan, Sarah Blackford

- Ten Simple Rules for Developing a Short Bioinformatics Training Course

  October 27, 2011 Allegra Via, Javier De Las Rivas, Teresa Attwood, David Landsman, Michelle Brazas, Jack Leunissen, Anna Tramontano, Maria Victoria Schneider

- Ten Simple Rules for Teaching Bioinformatics at the High School Level

  October 27, 2011 David Form, Fran Lewitter

- Ten Simple Rules To Combine Teaching and Research

  April 24, 2009 Quentin Vicens, Philip E. Bourne

- Ten Simple Rules for Graduate Students

  November 30, 2007 Jenny Gu, Philip E Bourne

## Event Planning

- Ten Simple Rules for organizing a non–real-time web conference

  March 26, 2020 Ana Arnal, Pablo Gregori, Vicente Martínez, Irene Epifanio

- Ten simple rules for organizing a webinar series

  April 1, 2019 Faisal M. Fadlelmola, Sumir Panji, Azza E. Ahmed, Amel Ghouila, Wisdom A. Akurugu, Jean-Baka Domelevo Entfellner, Oussema Souiai, Nicola Mulder, H3ABioNet Research working group as members of the H3Africa Consortium

- Ten simple rules for measuring the impact of workshops

  August 30, 2018 Shoaib Sufi, Aleksandra Nenadic, Raniere Silva, Beth Duckles, Iveta Simera, Jennifer A. de Beyer, Caroline Struthers, Terhi Nurmikko-Fuller, Louisa Bellis, Wadud Miah, Adriana Wilde, Iain Emsley, Olivier Philippe, Melissa Balzano, Sara Coelho, Heather Ford, Catherine Jones, Vanessa Higgins

- Ten Simple Rules on How to Organize a Scientific Retreat

  February 2, 2017 Julia Ponomarenko, Romina Garrido, Roderic Guigó

- Ten Simple Rules for a Community Computational Challenge

  April 23, 2015 Iddo Friedberg, Mark Wass, Sean Mooney, Predrag Radivojac

- Ten Simple Rules for Organizing an Unconference

  January 29, 2015 Aidan Budd, Holger Dinkel, Manuel Corpas, Jonathan Fuller, Laura Rubinat, Damien Devos, Pierre Khoueiry, Konrad Förstner, Fotis Georgatos, Francis Rowland, Malvika Sharan, Janos Binder, Tom Grace, Karyn Traphagen, Adam Gristwood, Natasha Wood

- Ten Simple Rules to Achieve Conference Speaker Gender Balance

  November 20, 2014 Jennifer Martin

- Ten Simple Rules for Running Interactive Workshops

  February 27, 2014 Katrina Pavelin, Sangya Pundir, Jennifer Cham

- Ten Simple Rules for Organizing a Virtual Conference—Anywhere

  February 26, 2010 Nelson N. Gichora, Segun A. Fatumo, Mtakai V. Ngara, Noura Chelbat, Kavisha Ramdayal, Kenneth B. Opap, Geoffrey H. Siwo, Marion O. Adebiyi, Amina El Gonnouni, Denis Zofou, Amal A. M. Maurady, Ezekiel F. Adebiyi, Etienne P. de Villiers, Daniel K. Masiga, Jeffrey W. Bizzaro, Prashanth Suravajhala, Sheila C. Ommeh, Winston Hide

- Ten Simple Rules for Chairing a Scientific Session

  September 25, 2009 Alex Bateman, Philip E. Bourne

- Ten Simple Rules for Organizing a Scientific Meeting

  June 27, 2008 Manuel Corpas, Nils Gehlenborg, Sarath Chandra Janga, Philip E. Bourne

## Practicing More Disciplined and Organized Science

- Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks

  July 25, 2019 Adam Rule, Amanda Birmingham, Cristal Zuniga, Ilkay Altintas, Shih-Cheng Huang, Rob Knight, Niema Moshiri, Mai H. Nguyen, Sara Brin Rosenthal, Fernando Pérez, Peter W. Rose

- Ten simple rules for carrying out and writing meta-analyses

  May 16, 2019 Diego A. Forero, Sandra Lopez-Leon, Yeimy González-Giraldo, Pantelis G. Bagos

- Ten simple rules on how to create open access and reproducible molecular simulations of biological systems

  January 17, 2019 Arne Elofsson, Berk Hess, Erik Lindahl, Alexey Onufriev, David van der Spoel, Anders Wallqvist

- Ten simple rules for responsible big data research

  March 30, 2017 Matthew Zook, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, Frank Pasquale

- Ten Simple Rules for Digital Data Storage

  October 20, 2016 Edmund M. Hart, Pauline Barmby, David LeBauer, François Michonneau, Sarah Mount, Patrick Mulrooney, Timothée Poisot, Kara H. Woo, Naupaka B. Zimmerman, Jeffrey W. Hollister

- Ten Simple Rules for Effective Statistical Practice

  June 9, 2016 Robert E. Kass, Brian S. Caffo, Marie Davidian, Xiao-Li Meng, Bin Yu, Nancy Reid

- Ten Simple Rules for Selecting a Bio-ontology

  February 11, 2016 James Malone, Robert Stevens, Simon Jupp, Tom Hancocks, Helen Parkinson, Cath Brooksbank

- Ten Simple Rules for Creating a Good Data Management Plan

  October 22, 2015 William K. Michener

- Ten Simple Rules for Experiments' Provenance

  October 20, 2015 Toni Kazic

- Ten Simple Rules for Protecting Research Integrity

  October 1, 2015 David Shaw, Thomas Erren

- Ten Simple Rules for the Care and Feeding of Scientific Data

  April 24, 2014 Alyssa Goodman, Alexander W. Blocker, Christine L. Borgman, Kyle Cranmer, Merce Crosas, Rosanne Di Stefano, Yolanda Gil, Paul Groth, Margaret Hedstrom, David W. Hogg, Vinay Kashyap, Ashish Mahabal, Aneta Siemiginowska, Aleksandra Slavkovic, Alberto Pepe

- Ten Simple Rules for Effective Computational Research

  March 27, 2014 James Osborne, Miguel Bernabeu, Maria Bruna, Ben Calderhead, Jonathan Cooper, Neil Dalchau, Sara-Jane Dunn, Alexander Fletcher, Robin Freeman, Derek Groen, Bernhard Knapp, Greg McInerny, Gary Mirams, Joe Pitt-Francis, Biswa Sengupta, David Wright, Christian Yates, David Gavaghan, Stephen Emmott, Charlotte Deanne

- Ten Simple Rules for Reproducible Computational Research

  October 24, 2013 Geir Sandve, Anton Nekrutenko, James Taylor, Eivind Hovig

- Ten Simple Rules for Doing Your Best Research, According to Hamming

  October 26, 2007 Thomas C Erren, Paul Cullen, Michael Erren, Philip E Bourne

## Programming and Software Management

- Ten simple rules for documenting scientific software

  December 20, 2018 Benjamin Lee

- Ten simple rules for biologists learning to program

  January 4, 2018 Maureen A. Carey, Jason A. Papin

- Ten simple rules for making research software more robust

  April 13, 2017 Morgan Taschuk, Greg Wilson

- Ten Simple Rules for Developing Usable Software in Computational Biology

  January 5, 2017 Markus List, Peter Ebert, Felipe Albrecht

- Ten Simple Rules for Developing Public Biological Databases

  November 10, 2016 Mohamed Helmy, Alexander Crits-Christoph, Gary Bader

- Ten Simple Rules for Taking Advantage of Git and GitHub

  July 14, 2016 Yasset Perez-Riverol, Laurent Gatto, Rui Wang, Timo Sachsenberg, Julian Uszkoreit, Felipe da Veiga Leprevost, Christian Fufezan, Tobias Ternent, Stephen J. Eglen, Daniel S. Katz, Tom J. Pollard, Alexander Konovalov, Robert M. Flight, Kai Blin, Juan Antonio Vizcaíno

- Ten Simple Rules for the Open Development of Scientific Software

  December 6, 2012 Andreas Prlić, James B. Procter

## Scientific Communication

- Ten simple rules for innovative dissemination of research

  April 16, 2020 Tony Ross-Hellauer, Jonathan P. Tennant, Viltė Banelytė, Edit Gorogh, Daniela Luzi, Peter Kraker, Lucio Pisacane, Roberta Ruggieri, Electra Sifacaki, Michela Vignoli

- Ten simple rules for getting started on Twitter as a scientist

  February 10, 2020 Veronika Cheplygina, Casper Albers, Natalia Bielczyk, Ionica Smeets, Felienne Hermans

- Ten simple rules to create biological network figures for communication

  September 26, 2019 G. Elisabeta Marai, Bruno Pinaud, Katja Bühler, Alexander Lex, John H. Morris

- Ten simple rules for writing statistical book reviews

  January 24, 2019 Christopher J. Lortie

- Ten simple rules for collaboratively writing a multi-authored paper

  November 15, 2018 Marieke A. Frassl, David P. Hamilton, Blaize A. Denfeld, Elvira de Eyto, Stephanie E. Hampton, Philipp S. Keller, Sapna Sharma, Abigail S. L. Lewis, Gesa A. Weyhenmeyer, Catherine M. O'Reilly, Mary E. Lofton, Núria Catalán

- Ten simple rules for scientists: Improving your writing productivity

  October 4, 2018 Todd C. Peterson, Sofie R. Kleppner, Crystal M. Botham

- Ten simple rules for responsible referencing

  April 12, 2018 Bart Penders

- Ten simple rules for writing a popular science book

  February 1, 2018 Adam J. Kucharski

- Ten simple rules for drawing scientific comics

  January 4, 2018 Jason E. McDermott, Matthew Partridge, Yana Bromberg

- Ten simple rules for writing a response to reviewers

  October 12, 2017 William Stafford Noble

- Ten simple rules for structuring papers

  September 28, 2017 Brett Mensh, Konrad Kording

- Ten simple rules to consider regarding preprint submission

  May 4, 2017 Philip Bourne, Jessica Polka, Ronald Vale, Robert Kiley
- Ten simple rules for short and swift presentations

  March 30, 2017 Christopher J. Lortie
- Ten Simple Rules for Writing a Reply Paper

  October 8, 2015 Mark Simmons
- Ten Simple (Empirical) Rules for Writing Science

  April 30, 2015 Cody Weinberger, James Evans, Stefano Allesina
- Ten Simple Rules for Effective Online Outreach

  April 16, 2015 Holly M. Bik, Alistair D. M. Dove, Miriam C. Goldstein, Rebecca R. Helm, Rick MacPherson, Kim Martini, Alexandria Warneke, Craig McClain
- Ten Simple Rules for Writing a PLOS Ten Simple Rules Article

  October 23, 2014 Harriet Dashnow, Andrew Lonsdale, Philip Bourne
- Ten Simple Rules for Better Figures

  September 11, 2014 Nicolas Rougier, Michael Droettboom, Bourne Philip
- Ten Simple Rules of Live Tweeting at Scientific Conferences

  August 21, 2014 Sean Ekins, Ethan Perlstein
- Ten Simple Rules for Writing Research Papers

  January 30, 2014 Weixiong Zhang
- Ten Simple Rules for Writing a Literature Review

  July 18, 2013 Marco Pautasso
- Ten Simple Rules for Editing Wikipedia

  September 30, 2010 Darren W. Logan, Massimo Sandal, Paul P. Gardner, Magnus Manske, Alex Bateman
- Ten Simple Rules for a Good Poster Presentation

  May 25, 2007 Thomas C Erren, Philip E Bourne
- Ten Simple Rules for Making Good Oral Presentations

  April 27, 2007 Philip E Bourne
- Ten Simple Rules for Getting Published

  October 28, 2005 Philip E Bourne

## Self Study/Learning Habits

- Ten Simple Rules for Lifelong Learning, According to Hamming

  February 26, 2015 Thomas Erren, Tracy Slanger, J. Groß, Philip Bourne, Paul Cullen
- Ten Simple Rules for Online Learning

  September 13, 2012 David Searls

## Service

- Ten simple rules for providing effective bioinformatics research support

  March 26, 2020 Judit Kumuthini, Michael Chimenti, Sven Nahnsen, Alexander Peltzer, Rebone Meraba, Ross McFadyen, Gordon Wells, Deanne Taylor, Mark Maienschein-Cline, Jian-Liang Li, Jyothi Thimmapuram, Radha Murthy-Karuturi, Lyndon Zass

- Ten simple rules for providing optimal administrative support to research teams

  October 3, 2019 Romina Garrido, Casandra A. Trowbridge, Nana Tamura

- Ten simple rules to initiate and run a postdoctoral association

  August 17, 2017 Chiara Bruckmann, Endre Sebestyén

- Ten simple rules for forming a scientific professional society

  March 23, 2017 Bruno A. Gaëta, Javier De Las Rivas, Paul Horton, Pieter Meysman, Nicola Mulder, Paolo Romano, Lonnie Welch

- Ten Simple Rules for a Bioinformatics Journal Club

  January 28, 2016 Andrew Lonsdale, Jocelyn Sietsma Penington, Timothy Rice, Michael Walker, Harriet Dashnow

- Ten Simple Rules for Starting a Regional Student Group

  November 21, 2013 Avinash Shanmugan, Geoff Macintyre, Magali Michaut, Thomas Abeel

- Ten Simple Rules for Providing a Scientific Web Resource

  May 26, 2011 Sebastian J. Schultheiss

- Ten Simple Rules for Reviewers

  September 29, 2006 Philip E Bourne, Alon Korngreen

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# One thousand simple rules

**Philip E. Bourne**[1]*, **Fran Lewitter**[2¶], **Scott Markel**[3], **Jason A. Papin**[4]

**1** Data Science Institute, University of Virginia, Charlottesville, VA, United States of America, **2** Bioinformatics and Research Computing, Whitehead Institute, Cambridge, MA, United States of America, **3** Dassault Systèmes BIOVIA, San Diego, CA, United States of America, **4** Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, United States of America

¶ Retired
* peb6a@virginia.edu

What began as a one-off in 2005 as Ten Simple Rules for Getting Published [1] has, in thirteen years, now multiplied a hundredfold to become One Thousand Simple Rules for many aspects of one's professional development and led to Quick Tips in the journal's Education section. This milestone of a thousand rules has been reached thanks to the unselfish work of all stakeholders—authors, editors, reviewers, and readers. Let's face it, writing, editing or reviewing a Ten Simple Rules (TSR) article is not the same as a publication that advances a scientific field. What it is going to get you is the satisfaction of knowing you have passed on a part of your experience in a form that is easily understood and acted upon by those following in your footsteps and hence have a very different kind of positive impact on science. Thank you. On the other hand, as a reader the TSRs might contribute in some small way to you getting tenure, or whatever else you care about in your professional life.

> *Insight for impactful publications. Thank you @PLOSCompBiol, this is great for my students as they start writing*!
>
> @jkellogg916

These are our words. In the summer of 2018 PLOS sent out a survey to both the authors and readers of TSRs to solicit your thoughts on the series. A summary of some of your words is presented in Table 1 and Table 2.

The story of how the TSRs started was outlined in that first article [1]. Suffice it to say here that the articles proved then, as now, a succinct and easily digestible piece of advice. Qualitative measures of success of the TSRs appear as framed versions on the walls in labs, or in the case of one of us (PEB) a stand in the corridor outside the lab which contained the collection of TSRs which periodically needed to be replenished. The fact that the stand was located near the bathroom and was eventually stolen will be left unexplored. The definitive qualitative measure of success came with the "Ten Simple Rules for Writing Ten Simple Rules" [2].

**Table 1. What readers are saying about TSR that influenced them in their career.**

| |
|---|
| *"It is difficult to choose one particular article as a favorite. I would like to point out that, as I progressed from being a Master's student doing research to a more experienced graduate student nearing graduation, a different set of 10SR articles have guided/influenced me."* Sarvesh Nikumbh, Graduate student |
| *"How to write a review, by M Pautasso. It changed my approach to writing reviews."* Anonymous, Academic faculty |
| *"Ten Simple Rules for Reproducible Computational Research, and other articles related to computational biology, programming, etc. They really helped in organizing my projects and code more efficiently."* Endre Sebestyén, Postdoc |

https://doi.org/10.1371/journal.pcbi.1006670.t001

OPEN ACCESS

**Citation:** Bourne PE, Lewitter F, Markel S, Papin JA (2018) One thousand simple rules. PLoS Comput Biol 14(12): e1006670. https://doi.org/10.1371/journal.pcbi.1006670

**Published:** December 20, 2018



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

**Table 2. What authors are saying about writing a TSR article.**

| |
|---|
| *"It led to the collaboration with a sociologist (J. Evans), and was highly discussed on social media, increasing the visibility of my laboratory."* Stefano Allesina |
| *"I want to continue developing ways to support undergraduate research because I feel that researchers don't effectively use or train undergrads and take them for granted."* Ben Harris |
| *"We simply wanted to share our experience in setting up a postdoc association, and discuss postdoc life in general. Also, this seemed like a good opportunity to have an "official track record" in leadership skills as a postdoc."* Endre Sebestyén |
| *"After having read other TSR that I found really useful. I wanted to take the challenge of synthesizing my ideas on the topic I chose (better figures)."* Nicolas Rougier |
| *"Writing our article (Ten Simple Rules for biologists learning to program) helped me build confidence in myself as a mentor and teacher."* Maureen Carey |

https://doi.org/10.1371/journal.pcbi.1006670.t002

A more quantitative measure of success comes from the growth of the series (Fig 1) and the article level metrics (ALMs) that PLOS keeps on each article. As of October 2018, there have been a combined total of 8.3 million views and downloads of the TSRs from the PLOS and PubMed-Central (PMC) Websites. Twelve articles each have over 100,000 views and downloads with "Ten Simple Rules for Writing a Literature Review" [3] topping the list with a staggering 1.1 million combined views and downloads. While not consistently translated, various articles have been made available in at least Spanish, Japanese, Russian, Norwegian, Chinese, and Farsi.

> *I'm addicted to Ten simple rules in @PLOSCompBiol ! #reading how to making research software bulletproof*
>
> @TivadarDanka

It's hard to capture the breadth of what has been covered in these one thousand rules (Fig 2). Perhaps the best way is with Ten Rules, well actually ten ways to subclass the one hundred articles. Those divisions are:

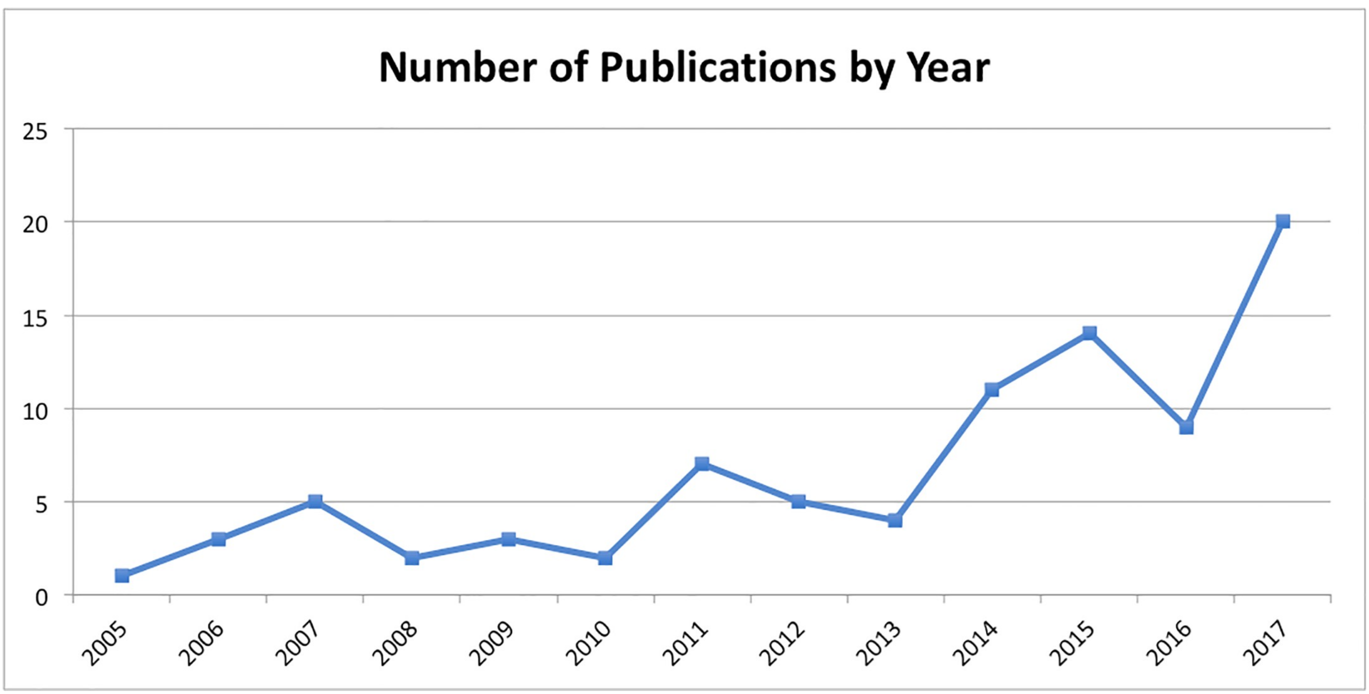



**Fig 1. Number of TSR articles published per year.**

https://doi.org/10.1371/journal.pcbi.1006670.g001

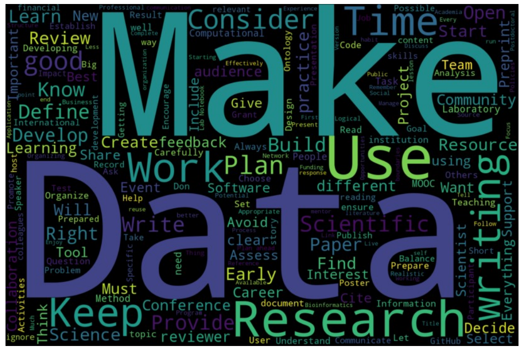



Fig 2. **Word cloud derived from the text of TSRs.** Courtesy of Andrew Lonsdale, November 2018.

https://doi.org/10.1371/journal.pcbi.1006670.g002

1. Self study/learning habits
2. Scientific communication
3. Career development and choices
4. Event planning
5. Education and mentoring
6. Using technology
7. Programming and software management
8. Service
9. Collaboration from local to international
10. Practicing more disciplined and organized science

With this level of coverage, has the series run its course? Is there nothing new to say? A question we have asked at PLOS a number of times only to have it answered by a new and novel submission. Certainly with one thousand rules novelty becomes an issue. Then again, in just reading responses to the reader survey which asked for novel TSR suggestions new ideas emerged—unique features of working for an NGO such as a non-profit foundation; reducing your carbon footprint while doing research; lab management; ensuring a work-life balance—and so on. Moreover, technologies, protocols, expectations, how we conduct science—think social media—are changing rapidly which calls for topics to be refreshed. In short, there would seem to be no end in sight.

There are also the questions of scope and form of dissemination. Many of the rules are not confined to computational biology and are clearly read by those outside our field. Since 2005 social media has become mainstream and while articles get tweeted are they more appropriate as blog postings? For now, we (PLOS and the TSR Editors), based on comments received through the surveys, feel we should keep the series going as is, but as a community journal your ongoing input is critical. What do you think comes next?

## References

1. Bourne PE. (2005) Ten simple rules for getting published. PLoS Comput Biol 2005; 1(5): e057. https://doi.org/10.1371/journal.pcbi.0010057 PMID: 16261197.

2. Dashnow H, Lonsdale A, Bourne PE (2014) Ten Simple Rules for Writing a PLOS Ten Simple Rules Article. PLoS Comput Biol 10(10): e1003858. https://doi.org/10.1371/journal.pcbi.1003858. PMID: 25340653

3. Pautasso M (2013) Ten Simple Rules for Writing a Literature Review. PLoS Comput Biol 9(7): e1003149. https://doi.org/10.1371/journal.pcbi.1003149. PMID: 23874189

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for more objective decision-making

**Anthony C. Fletcher[1], Georges A. Wagner[2], Philip E. Bourne[3]***

**1** Retired, London, United Kingdom, **2** Department of Mechanical Engineering, University of Melbourne, Victoria, Australia, **3** School of Data Science, University of Virginia, Charlottesville, Virginia, United States of America

* peb6a@virginia.edu





**Citation:** Fletcher AC, Wagner GA, Bourne PE (2020) Ten simple rules for more objective decision-making. PLoS Comput Biol 16(4): e1007706. https://doi.org/10.1371/journal.pcbi.1007706

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** April 2, 2020



**Funding:** No sponsors or funders for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Scientists spend their lives analyzing data by the systematic study of the structure and behavior of the physical and natural world using both observation and experiment—objective analysis. But when it comes to decision-making, scientists are also humans with accompanying subjectivity. Put colloquially, we have both heart and head—and are capable of being simultaneously subjective and objective.

Here we posit that bringing more objectivity ("head") to decisions is a good thing. It's a key part of "critical thinking," the "Socratic questioning" method. We are not suggesting, that like Mr. Spock, we should be driven entirely by rationality, nor are we considering the merits of various reasoning systems [1]; we are simply examining why greater objectivity helps in providing a simple way to achieve improved objectivity. So, to start, is objectivity indeed better than subjectivity?

To address this question, it's useful to look at the 2 opposite ends of the spectrum: objectivity is really the application of pure logic (something is either right or wrong, more or less, etc.), whereas subjectivity [2] is embodied in the form of what is often called Cartesian Doubt or skepticism (that the knowledge of anything outside ones direct experience has to be considered as unsure). In certain cases, increased objectivity is superior, for example, when the decision being taken leads toward a measurable or quantifiable outcome: if there is a specific goal in mind, then it's very useful to be able to estimate how close that decision might get you to that goal before you set out on the path. In real life, most decisions are a mixture of head and heart, but with these rules, we hope to increase both the accuracy and quantity of the head part while not neglecting the heart.

But enough of the epistemological concepts, what we want is to make better decisions (better here being more objective) and look at 10 ways in which we might do this, culminating in a simple tool that anyone with a spreadsheet (or even a pen and paper) can use.

Each rule is accompanied by a use case, some drawn from 2 previous Ten simple rules: Ten simple rules for graduate students [3] and Ten simple rules for selecting a postdoctoral position [4].

We will culminate with a worked example that illustrates this approach. Every lab needs a good coffee machine, and we are inspired by the example of the famous Trojan Room coffee pot. Based in the old computer laboratory of the University of Cambridge, England, in 1991, it provided the inspiration for the world's first webcam [5]. So here we show how to make sure your coffee is well up to par!

## Rule 1: Break decisions into smaller parts

Most decisions involve a range of issues that all need to be weighed up, not only in a like-for-like comparison, but also considering nonsimilar elements. The idiom about comparing apples to oranges serves this well: yes, they are clearly different, but they also have lots of similarities (you can compare the calorific content, the "mouth feel", nutrient make-up, etc.), Ultimately, you might like the taste of one but not the other (which is essentially subjective) or you might need to stave off scurvy (which is objective), but all things being equal, you could compare the characteristics and see whether one is preferable to the other under a given situation.

Thus, we can break down the apple versus orange decision into smaller parts (elements) as fully objective comparisons, subjective comparisons, and some that might be termed quasi-objective, in that they can rely on third party or expert opinion being nonpartisan and at least derived, in part, from a desire to be data-driven. Example studies comparing apples and oranges have been published by both NASA [6] and in the BMJ [7].

An early pioneer of this approach was Ramon Llull, a medieval Franciscan monk who postulated that if you had a series of basic "truths," these could be combined to make greater sets of truths that were, de facto, universally acceptable—a form of combinatorics. As a footnote, it's important to state that problems can be cleanly decomposed or factored in this way only if the subproblems are independent of each other.

### Case study

I've finished my undergraduate degree and am looking at where to study my PhD.

Often this decision is expressed as: should I study A at X or B at Y. Clearly, it's very complex to compare these 2 options. In order to not to miss possible solutions, we must avoid considering any aspect in isolation. Thus, although a valid approach is to consider the merits of A over B (if course content is our primary decider), this leads to a de facto answer of institution X or Y. Equally considering X over Y first leads to either A or B. Thus, we have to include all the characteristics of both the course and the institution in the decision matrix.

## Rule 2: Mitigate bias with the right set of inputs

Confirmation bias is a tendency to search for or interpret information in a way that confirms existing preconceptions. This means you put more weight on things you implicitly agree with and less on those that appear to go against your initial main thrust of a "good" decision. It is easy to unbalance the whole process if you start off with a predetermined idea and ignore key elements. Guarding against this is very complex and the subject of much academic study.

In practice, assembling and measuring the objective elements of decisions is usually straightforward—you can look up data—but you have to be measuring the right sets of elements. Most areas have, at the very least, quasi-objective data easily available, e.g., professional or reputable reviewers, specialist journalists, etc., and it's worth looking to see that the elements that third parties might commonly use to appraise topics are also in your own review.

### Case study

I always thought that a top university is the place to study—is that true?

Many worldwide university rankings would support this opinion, or even have given rise to it. However, academic cultures vary from nation to nation and so do the teaching approaches. To take just one example, French scientific institutions tend to produce graduates with a very strong grounding in mathematics and these graduates often go on to excel in research departments worldwide. Therefore, it is a good idea to dig deeper than a university's general ranking when choosing which institution to attend. This means finding out what the reputation of a particular course or research group of interest is. Especially in research, the reputation of a particular lab may well differ starkly to the department's or even the university's reputation as a whole.

## Rule 3: Greater transparency helps in making the right decisions

Most decisions are influenced by others even if they aren't made with direct external input. By breaking down your decision into components (a number of which may be motivations) and sharing that breakdown, it's simpler to demonstrate reasoning, and it can then be more straightforward to persuade others your decision is the right one.

### Case study

This is really a restatement of the idea most of us became familiar with early on: it's not enough to get the right answer, you must also show your methods.

By exposing the workings, you avoid a "black box," and this can be incredibly useful when asking others to check your work for errors or simply asking them to follow your logic. For example, if you are administering staff and want to introduce a new workplace rule, explain why during consultations—that way, people are more amenable to accepting the underlying logic, and it can make it easier to negotiate problems as you can focus on resolving any minor issues without derailing the whole idea.

## Rule 4: Feedback loops should be utilized to improve future decision-making and inform others

By breaking down and codifying your reasoning, you can come back to it after the decision has played out. Did I make the right choices? And if not, what did I assume that was wrong in the original reasoning. Hindsight is wonderful if applied properly.

We make decisions with a surprising variation in frequency—trivial ones are made daily (what shall I have for lunch?), others yearly (where shall we vacation?), every several years (let's replace the car), or a few times in life (which property shall we buy; shall I take this job?). Many of these are repeated, so it's good not to repeat obvious errors. A few decisions are unique of course (what shall I study at university?).

But even trivial decisions are worth considering more closely. If you want to eat healthily, it's worth thinking objectively about your lunch habits, and if you were only partly satisfied with your last car, why was that? You can't do much about your choice of subject to study at university, but you can tell others about the results of your choice, and that helps inform them.

### Case study

For my PhD, I chose an emerging rather than an established field. How well did that work for me? If it worked well, I might consider the same approach for a postdoctoral position. If the risk paid off (or not), the experience should be shared.

## Rule 5: Get multiple opinions

In Rule 2, we considered bias. Multiple diverse expert opinions tend to mitigate bias. Do not be reticent in getting those opinions—your future may depend on it. Once obtained, consider all opinions carefully. A key word in this short rule is "expert"—on the internet everyone has an opinion, and of course, that is valid for them. But it might well not be very objective. So, take some time to consider the experience and qualifications of the source of the opinion.

### Case study

You are considering a field of study for your PhD, let's say, cell signaling. If you ask a cell signaling researcher, they will inevitably tell you it is an admirable field of study—we tend to create students in our own image. So, ask other scientists from different fields for their view and weigh all advice carefully.

## Rule 6: Some decisions aren't yours to make alone

In the corporate and academic worlds, decisions are often made by groups, and we often bring to mind the old truism that "a camel is a horse designed by a committee." People involved in these decisions come at them with a varying degree of bias and subjectivity (for example, is this good for my department rather than the university as a whole). Again, by breaking decisions into discrete parts, in this case, per individual, such bias is more obvious, and it is more amenable to be settled objectively, and some level of consensus may emerge.

### Case study

Along with colleagues, you may have shortlisted 3 candidates for a tenure-track position. But what if they don't agree with your choice?

You will have to argue both the facts and perhaps the emotion. You can then separate the fact part of the discussion from the emotional side and treat these appropriately, perhaps by going through the decision process step-by-step.

## Rule 7: Beware of cognitive dissonance

Cognitive dissonance is the psychological stress experienced by a person who holds 2 or more contradictory beliefs, ideas, or values [8]. People experiencing such conflicts generally try to reduce the psychological discomfort by attempting to achieve emotional equilibrium in a variety of ways, for example, downplaying, outweighing, repressing, or incorporating information. When they put these beliefs under stress (for example, in decision-making), they can experience problems, especially if they receive new information that clashes with an existing belief or notion.

These issues all mitigate against good decision-making. But it's worthwhile noting that real dilemmas can occur and they are worth flagging if they do. It's an interesting side note to recognize that a lemma is a small truth, a stepping stone on the way to a larger truth (a theorem). So, a dilemma is just 2 opposite truths, and you can have a trilemma as well. Again, by breaking a decision down into component parts, we help to recognize the presence of such dilemmas and find ways to resolve them.

### Case study

You have a parent whose graduate studies were in A, and they would like you to follow in their footsteps. But field B really stands out, and this is why you prefer it.

If you simply rely on facts, you will never get to the emotional feelings being expressed by the parent, so you have to connect the 2. One good way to do this is to ask open-ended questions (ones that cannot be answered with a yes or no) and to argue the facts or to present a fact and ask the parent how this might make them feel. Emotions are real and have to be dealt with. Objectivity only goes so far.

## Rule 8: Not all systems are amenable to objective decisions

When a decision is tied to strong emotional elements, you have to decide if objectivity is the right way to satisfy yourself in the longer term. You will have to live with your choice and perhaps deal with the psychological repercussions. It's one thing in the cold light of day to decide A, B, and C are the things you most like about, say, a city to go and live or work in, but several years on, you may well wonder why you never felt really at home there.

### Case study

A typical example of this are human partnerships in which both people are pursuing equally important but different careers. For one party, the ideal place to work might be A, whereas for the other party it is B (A and B being in different cities or countries). There may be no objective answer to this.

## Rule 9: What if? Working backwards from a range of solutions

"What if" scenarios are in which a range of possibilities are modeled to look at possible outcomes. After a "base model" is produced, inputs are then varied, and outputs scrutinized. Weightings are then varied, and how this changes the decision can be reviewed. This gives you a very good idea about what is driving a decision in a certain way and may allow room to reflect further on your original choices of key important elements. However, it should be mentioned that blind "goal seeking" by manipulating inputs is generally a terrible way to make a choice and also that this is not a simple binary decision tree type of outcome (if this: then that).

### Case study

When thinking about a university for graduate studies, there are a set of options (outputs). Features include various ranking criteria, the prestige of the lab you are joining, the labs current productivity, etc. Weighting those factors by what is important to you will point you to a solution.

## Rule 10: Apply the weighted decision matrix

And so, to Rule 10—a way to easily increase objectivity in decision-making. For those interested in history, Pascal's Wager is often considered the first example of decision theory [9]. Nowadays, however, decision matrices will be a tool familiar to anyone who has come across Six Sigma (6σ), a set of techniques and tools for process improvement in industry [10], and the Pugh Matrix [11], a more generalized description of the process.

At their most basic, these matrices require you to break a decision down to a set of discrete elements or criteria and then compare these to each other, which gives a level of internal consistency. Weighting can then improve the results further.

There is a large body of literature on decision theory and choice under uncertainty, and although it is unrealistic to try to cover it in this short article, it should be mentioned that it has been a subject of study for centuries and has led to several Nobel prizes in areas such as expected utility theory and prospect theory.

### Use case

Every lab needs a good coffee machine, so let's use this simple example—you need a new "pod"-based coffee machine. We can start by comparing models from companies like Lavazza, Nespresso, Keurig, etc. We then break down the elements of these (for example, price of machine, price of pods, availability, quality of coffee, aesthetics of machine, etc.). Many of these elements are amenable to fact-based comparisons (price, for example), whereas others are amenable to what we might call metafacts (quality of coffee, where you can read reviews by coffee experts or simply try them yourselves), and others are essentially subjective (e.g., aesthetics). We can then assign values (e.g., from 0–10).

When this matrix is complete, we then add another column that applies a weighting: so, if you rate aesthetics more than coffee quality or don't care much about price, for example, you can reflect this. A simple calculation then provides a "winner." It's also very easy to do a "what if" on this by varying weightings. An example of this would be if I cared a bit more about the quality and little less about the price, or alternatively, it might be an ugly machine, but the coffee is great (Fig 1).

The range of uses for this type of analysis are wide, from deciding where to pursue your PhD or postdoc to deciding upon which house or car to buy.

The more the elements lend themselves to data, or metadata, the more useful they are, but it's not an exact tool. Subjectivity is being brought to many aspects, not least the weightings or even the selection of elements. However, because it's quick to do, it can provide useful pause for reflection when it comes to bigger decisions.

## Conclusion

These 10 rules consider, in a simplistic way, features and tools to help make decisions, career related or otherwise. We aren't arguing that fully objective decision-making is possible or even desirable. What we do argue is that a little time spent breaking down the elements of a decision leads to useful insights, and we outline a simple tool that is helpful in this process.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **weighted decision matrix example** | | | | | |
| 2 | | | **pod Type** | | | |
| 3 | | | **Nespressc** | **Lavazza** | **Dulce Gu:** | **Tassimo** |
| 4 | | **Weight** | | | | |
| 5 | quality of coffee | 9 | 7 | 8 | 6 | 6 |
| 6 | availability of pods | 6 | 5 | 6 | 8 | 7 |
| 7 | cost of machine | 7 | 4 | 7 | 8 | 8 |
| 8 | look of the machine | 3 | 9 | 7 | 5 | 7 |
| 9 | reliability/usefulness | 5 | 8 | 8 | 7 | 6 |
| 10 | | | | | | |
| 11 | **weighted score** | | **297** | **324** | **306** | **306** |
| 12 | | | | | | |
| 13 | | | | | | |

**Fig 1. A simple weighted matrix.**

https://doi.org/10.1371/journal.pcbi.1007706.g001

Abraham Lincoln was known to exercise objectivity by writing letters he never sent: the process of simply preparing an argument and committing it to paper is cathartic in itself, and it can also give unexpected insight into the issue at hand [12]. We simply extend that idea here.

Best of luck finding the best coffee machine for your lab.

## Acknowledgments

Thanks to Tim Collins and David Padmore.

## References

1. Goodmath. ScienceBlogs [Internet]. Basics: Logic, aka "It's illogical to call Mr. Spock logical." 2006 [cited 2020 Mar 15]. Available from: https://scienceblogs.com/goodmath/2007/01/27/basics-logic-aka-its-illogical-1
2. Solomon RC. Subjectivity. In: Honderich T, ed. Oxford Companion to Philosophy. Oxford: Oxford University Press; 2005. p. 900.
3. Gu J, Bourne PE. Ten simple rules for graduate students. PLoS Comput Biol. 2007; 3(11): e229. https://doi.org/10.1371/journal.pcbi.0030229 PMID: 18052537
4. Bourne PE, Friedberg I. Ten simple rules for selecting a postdoctoral position. PLoS Comput Biol. 2006; 2(11): e121. https://doi.org/10.1371/journal.pcbi.0020121 PMID: 17121457
5. BBC [Internet]. Witness History: The Trojan Room Coffee Pot. 2012 [cited 2020 Mar 15]. Available from: https://www.bbc.co.uk/programmes/p010lvn7
6. Sandford, SA. Improbable Research [Internet]. Apples and oranges—A Comparison. 1995 [cited 2020 Mar 15]. Available from: https://www.improbable.com/airchives/paperair/volume1/v1i3/air-1-3-apples.php
7. Barone JE. Comparing apples and oranges: a randomized prospective study. BMJ 2000. 321(7276): 1569–1570. https://doi.org/10.1136/bmj.321.7276.1569 PMID: 11124178
8. Cognitive dissonance [Internet]. Wikipedia [cited 2020 Mar 15]. Available from: https://en.wikipedia.org/wiki/Cognitive_dissonance
9. Pensees [Internet]. Wikipedia [cited 2020 Mar 15]. Available from: https://en.wikipedia.org/wiki/Pens%C3%A9es
10. Six sigma [Internet]. Wikipedia [cited 2020 Mar 15]. Available from: https://en.wikipedia.org/wiki/Six_Sigma
11. Decision-matrix method [Internet]. Wikipedia [cited 2020 Mar 15]. Available from: https://en.wikipedia.org/wiki/Decision-matrix_method
12. Konnikova M. The lost art of the unsent angry letter [Internet]. New York: New York Times. 2014 Mar 22 [cited 2020 Mar 15]. Available from: https://www.nytimes.com/2014/03/23/opinion/sunday/the-lost-art-of-the-unsent-angry-letter.html

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for landing on the right job after your PhD or postdoc

**Kuan-lin Huang** *

Department of Genetics and Genomic Sciences, Center for Transformative Disease Modeling, Tisch Cancer Institute, Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America

* kuan-lin.huang@mssm.edu

OPEN ACCESS

**Citation:** Huang K-l (2020) Ten Simple Rules for landing on the right job after your PhD or postdoc. PLoS Comput Biol 16(4): e1007723. https://doi.org/10.1371/journal.pcbi.1007723

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** April 2, 2020



**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Freshly minted PhD and postdocs can often benefit from thorough guidance on career development and choosing the right job afterward. Several articles in the Ten Simple Rules series help you navigate this challenge, especially on selecting a postdoctoral position [1], considering a career in academia versus government [2], choosing between industry and academia [3], starting a company [4], and approaching a new job [5]. While these articles mostly include invaluable advice from senior leaders reflecting from decades of experience, I (having just gone through the process) wish to offer a complementary set of fresh-baked lessons.

During my PhD and a brief postdoc, I started a company, consulted part time, and participated in science-policy and teaching group activities. I then interviewed at the academia, biotech, and pharmaceutical companies before deciding on my current tenure-track position at the Icahn School of Medicine at Mount Sinai at the end of 2018. I am fortunate to have obtained a wide range of experience and cultivated networks of people to learn from (albeit most are limited to the United States). Here, I distill the ten rules into three sections representing distinct phases of landing on the right job—exploration, decision, and fulfillment—with practical tips (companion video: https://youtu.be/O6HZJgqhxA4).

## Exploration

> *"Exploration is what you do when you don't know what you're doing. That's what scientists do every day."–Neil deGrasse Tyson*

### Rule 1: Know your values, goals, and priorities

Start by knowing yourself better. Life is a continuous journey of exploring and attaining one's purpose. What impact do you want to have on the world? Try Steven Covey's funeral exercise [6]: Picture attending your own funeral; what would you want your family, friends, colleagues, and others to say about you and what you have done? From the other angle, adopt Jim Collin's hedgehog concept [7] for a company for yourself: Find how you can best contribute from the intersect between (1) what you are deeply passionate about, (2) what you can be the best in the world at, and (3) what you can get paid for.

You can also achieve tangible answers by asking yourself specific questions. Are you more driven by autonomy or order? Fortune or fame? Knowledge or utility? How much do you value spending personal time outside of work for yourself and the family, creating new knowledge and exercising intellectual autonomy, or translating science into products that directly affect consumers' life? The other question that many people find helpful is as follows: If you were to pick one or a hybrid of multiple people as a model for your career, who would that be? What did that person do at the early stage of their career to help them become who they are?

At a practical level, reflect on what activity attracts or deters you from your work so far. Scientific research is often a job with multiple dimensions. You can dissect which aspects you particularly enjoy and thrive at. For example, do you enjoy the days when you are coding away or conducting experiments, or do you enjoy the days when you are writing manuscripts or drafting grants? Do you enjoy reading papers and planning experiments, or do you enjoy seeing the translational impacts that your research may have on others' lives? Do you enjoy constant personal interactions, or do you prefer moments of solitude? You may not have a dichotomized answer for each of these exploratory questions. But pay attention to your preference. You can make decisions to optimize for it.

## Rule 2: Talk to everyone and see the possibilities

Your view of the possibilities constrains your choice. Expanding this view can help you identify the global maximum of reaching your potential. Being in the academic environment, many of you default to your PhD and postdoc advisors or other professors for career advice. While faculty may provide helpful thoughts for an academic career, we often possess limited insights on diverse "real-world" opportunities, considering that many of us have been at schools for our whole lives. Make sure you consult with the career center at your institution to explore your options; many now provide counselors with PhDs. Take every opportunity you have in networking events to talk with the majority of the PhDs who choose a nonacademic career path and learn about it. Listen, keep an open mind, and stay in touch. People are generally happy to share about their careers and willing to help those who are interested in following similar paths. And down the road, a referral—even from a weak tie—can get your first foot in the door in job applications.

## Rule 3: Gather the data

As scientists, you show the data to validate hypotheses and inform decisions. It should come as second nature to apply this evidence-based approach to your career too. Read books and journal articles written by diverse professionals to learn about different fields and the emerging trends you may be excited about. Additional information is also increasingly available to help evaluate your options. For example, job sites like Glassdoor post employee reviews, salaries, and interview tips. These sources can help you gather useful information that is not common knowledge. For example, a critical managerial change may be documented in recent employee reviews. Cross-validating such reviews with what I hear from my personal network or during on-site interviews, I find that many of them provide reasonable honest opinions on hard-to-gauge, yet important, aspects like organizational cultures.

## Rule 4: Conduct "mini-experiments" to try it out

One of the best ways to find out what you like is to do it. Many research institutions now have special-interest graduate and postdoc groups that enable you to try-out different careers. These groups can help you explore teaching, consulting, entrepreneurship, science policy, and other opportunities with varying levels of commitments. Not sure whether a consulting job suits you? Sign up for a quarterly consulting project. Have an itch for starting a company? Join a nearby incubator and explore start-up projects with like-minded individuals. You can also explore opportunities intersecting science and other fields through classes, seminars, and events at other schools such as the business school or the law school. Aside from the hands-on experience, these opportunities can help you expand your network and find advice from people of diverse professions.

### Rule 5: Sympathize with your future self

Sympathy is an essential attribute of a socialized human being. When it comes to making career choices, using thought experiments to sympathize with your future self may reveal insights. For example, imagine you are a professional a few years or even decades further along on the path you chose. Would you be proud of what you have achieved? Are those choices aligned with your values? As another practical example, think about the day-to-day routine of each job that you have now learned from prior explorations. Imagine yourself going through a typical day. Do you enjoy how your time is allocated to different categories of activities, especially those that would preoccupy most of your days? For example, would you enjoy grant writing as a junior faculty? Market research as a consultant? Coding as a computational biologist? Stretch and compress your current allocations for similar tasks to the extent demanded by the future job, and you may be able to see and feel more clearly whether you will strive in it.

## Decision

*"Do nothing, and nothing happens. Life is about decisions. You either make them or they're made for you, but you can't avoid them."–Mhairi McFarlane*

### Rule 6: Evaluate the fit

Once you have gained a deeper understanding of yourself and the job, it is time to find whether the two align. Carefully examine the fit between your career goals, personality, and skillset and the organization's mission, culture, and job requirements. Do your values align with the organization's aims and goals? Carefully examine the organization's mission statement and, sometimes more telling, its core economy-driving activity. For example, many internet technology companies derive values by optimizing advertisement views and clicks, which drives the economy (with the potential trade-off on consumer privacy and mental health). As another example, within the same industry, different pharmaceutical companies target different diseases, serve different populations, and deploy different pricing schemes. Your job is designed to advance the organization's mission, which you will need to align to for long-term job fulfillment.

In evaluating personality fit, the unique challenges of the job often determine the organizational culture to a certain extent. For example, working in a large company likely requires you to execute excellence on your assigned task repeatedly. A job at a start-up probably involves a more dynamic role in which you will need to figure out the next challenges and solutions autonomously. A research-focused faculty job typically requires drafting of grant proposals relentlessly, especially in the early phase. Your role and the organization will evolve over time. Continue to evaluate, adjust, or—as needed—find a different fit.

### Rule 7. Get along with the five people around you

While each person is unique, it is natural to become more like the five people around you. The importance of getting to know your potential future teammates cannot be over emphasized. Most on-site interviews will arrange for you to meet with your prospective supervisors and colleagues. Make sure you interview them while they interview you. Are they happy? Can you imagine getting along with them? Do you share the same aspirations and levels of ambition? Would you enjoy a good time with them during work lunch and happy hours? If you could not gather enough information during the interview, follow up and politely ask whether you can ask more questions in a call or email.

### Rule 8: Plan out the dates for your transitions

Up until your PhD defense, all the start and end dates of school programs are set for you. You need to make the conscious switch that most future transition timings are jointly determined by you and the organization. A faculty job search can typically take a long duration with multiple on-site interviews. Industry and postdoc jobs tend to have faster turnarounds, and some may have strict deadlines for you to decide by. If your doctoral advisor holds conflicts in keeping you in the lab for longer, you may seek support from third-party faculty to determine if you have met the criteria of attaining a doctoral degree. If you are an international scholar, communicate early on with the international scholar office in your institution to make sure that your transition timing is aligned perfectly with your visas to avoid immigration issues. You need to take responsibility to align the dates for your unique circumstances.

### Rule 9: Don't forget: Location, location, location

Now you have considered the organization, team, and timing, one often under-emphasized factor is the location. A job comes with the location at which you will likely live for at least several years of your life. It is where you will build your network and thus have a higher likelihood of staying long term. Would you prefer the cost-effective and spacious living at suburban locations or the dynamic lifestyles and job-hopping opportunities in big cities? How about places with warm sunshine year-round versus those with defined seasons? If you are moving with your family, what do they prefer? Each location is unique. I highly recommend that you spend extra time during on-site interviews to gain a brief living experience.

## Fulfillment

> *"'How does one become a butterfly?' she asked pensively.*
>
> *'You must want to fly so much that you are willing to give up being a caterpillar.'"–Trina Paulus*

### Rule 10: Settle in, contribute, and keep learning

Transition is hard but also exciting. In the beginning, it is perfectly normal to be missing your previous lab and colleagues. Acknowledge that this is a natural process that most people go through and appreciate aspects of your new role and environment. There may also be times where you uncontrollably fantasize about the opportunities that you did not take. Identifying role models may provide some guiding lights—you can find plenty of accomplished individuals who stay with their grit and excel in one profession and just as many that switch tracks and contribute in multiple fields.

In the long term, finding meaning in your work is critical to a fulfilled life. Through knowing your values, goals, and priorities (Rule 1) and carefully evaluating the fit (Rule 6), you should be happily contributing to your and the organization's shared mission with the right people in the right place (thanks to Rule 5, 7, 9). While the progress may have ups and downs, remember it took all your PhD and postdoc years to make a dent in expanding human knowledge. Given the right alignment, what you do every day shall accumulate to a substantial impact. Remember that landing on the job is part of a journey, not an end. You can always apply these rules to assess your situation and change courses.

Finally, as a scientist, you know that there is always plenty more to learn. Endowed with the knowledge and determination of a PhD, you are in a great position to contribute. But remember to maintain your gift of curiosity. Seek opportunities to grow. Continue to solve new challenges that are important to you and society. After all, a fulfilling career is not unlike a PhD—you ask an important question, experiment to solve it, and pursue the next one.

## References

1. Bourne PE, Friedberg I. Ten simple rules for selecting a postdoctoral position. PLoS Comput Biol. 2006. https://doi.org/10.1371/journal.pcbi.0020121 PMID: 17121457
2. Bourne PE. Ten simple rules in considering a career in academia versus government. PLoS Comput Biol. 2017. https://doi.org/10.1371/journal.pcbi.1005729 PMID: 28981499
3. Searls DB. Ten simple rules for choosing between industry and academia. PLoS Comput Biol. 2009. https://doi.org/10.1371/journal.pcbi.1000388 PMID: 19668326
4. Fletcher AC, Bourne PE. Ten simple rules for starting a company. PLoS Comput Biol. 2012. https://doi.org/10.1371/journal.pcbi.1002439 PMID: 22479171
5. Bourne PE. Ten Simple Rules for Approaching a New Job. PLoS Comput Biol. 2014. https://doi.org/10.1371/journal.pcbi.1003660 PMID: 24967974
6. Covey SR. The Seven Habits of Highly Effective People: Restoring the Character Ethic. Simon and Schuster; 1989.
7. Collins JC. Good to Great. HarperBusiness; 2001.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules to becoming a principal investigator

**John S. Tregoning**[1]*, **Jason E. McDermott**[2,3]*

**1** Department of Infectious Disease, St Mary's Campus, Imperial College London, London, United Kingdom, **2** Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, United States of America, **3** Department of Molecular Microbiology and Immunology, Oregon Health & Sciences University, Portland, Oregon, United States of America

* john.tregoning@imperial.ac.uk (JST); Jason.McDermott@pnnl.gov (JEM)



OPEN ACCESS

**Citation:** Tregoning JS, McDermott JE (2020) Ten Simple Rules to becoming a principal investigator. PLoS Comput Biol 16(2): e1007448. https://doi.org/10.1371/journal.pcbi.1007448

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** February 20, 2020



**Funding:** The authors received no specific funding for this study.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

The biggest choke point in an academic career is going from postdoc to principal investigator (PI): moving from doing someone else's research to getting other people to do yours. Being a PI is a fundamentally different job to being a postdoc; they just happen to be in the same environment. It is not an easy transition. It draws on few of the skills you learn at the bench, and the odds are clearly not ever in your favor. So, calling this article Ten Simple Rules is obviously a simplification. It is more accurate to call them ten tricky steps.

In this article, we use PI to mean anyone who runs their own research group using funding that they have been awarded to answer their own questions. PI encompasses a number of different job titles depending on where the research is performed: fellow, lecturer, reader, associate professor, and senior scientist. One test is whether you can describe the people working for you as the X group, in which X is your surname. The normal route from undergraduate to lab head involves a PhD, one or more postdoc positions, and then PI. Given the diversity of ways to be a PI, the final step up from postdoc takes a number of forms. In the United Kingdom, this tends to be either an individual fellowship or a lecturer position, and in the United States, it generally starts with an independent position with associated funding—either as a start-up package or funded grant.

The aim of this article is to identify some of the broader skills (rules 1–4) and behaviors (rules 5–10) that can help with getting a PI position. It is meant as advice not instruction. As you will see, we are advocating the development of social intelligence, which is as useful in the world outside academia as within it.

## Rule 1: Have ideas

Creativity is central to being a PI—seeing new connections, thinking of new ideas, and using current understanding to develop future plans. Unfortunately, creativity is incredibly nebulous and can feel at odds to the scientific process (Fig 1). Be receptive to ideas whenever they come, especially as they often come at the most inconvenient of times—when dropping the kids off or at 4:00 in the morning. Find ways to capture these flitting ideas. Accept that there are few truly novel ideas: Reading around will provide you with inspiration for your own problems. Whilst it is more about making things than doing science, *Every Tool's A Hammer* captures what it is to be creative [1].

Learn to accept that ideas do not just come by themselves and allow them time to develop: It is entirely normal to have more bad ideas than good ones. Even if the net product from the day is a waste bin full of paper and some tea-stained scribbles—having a creative process,

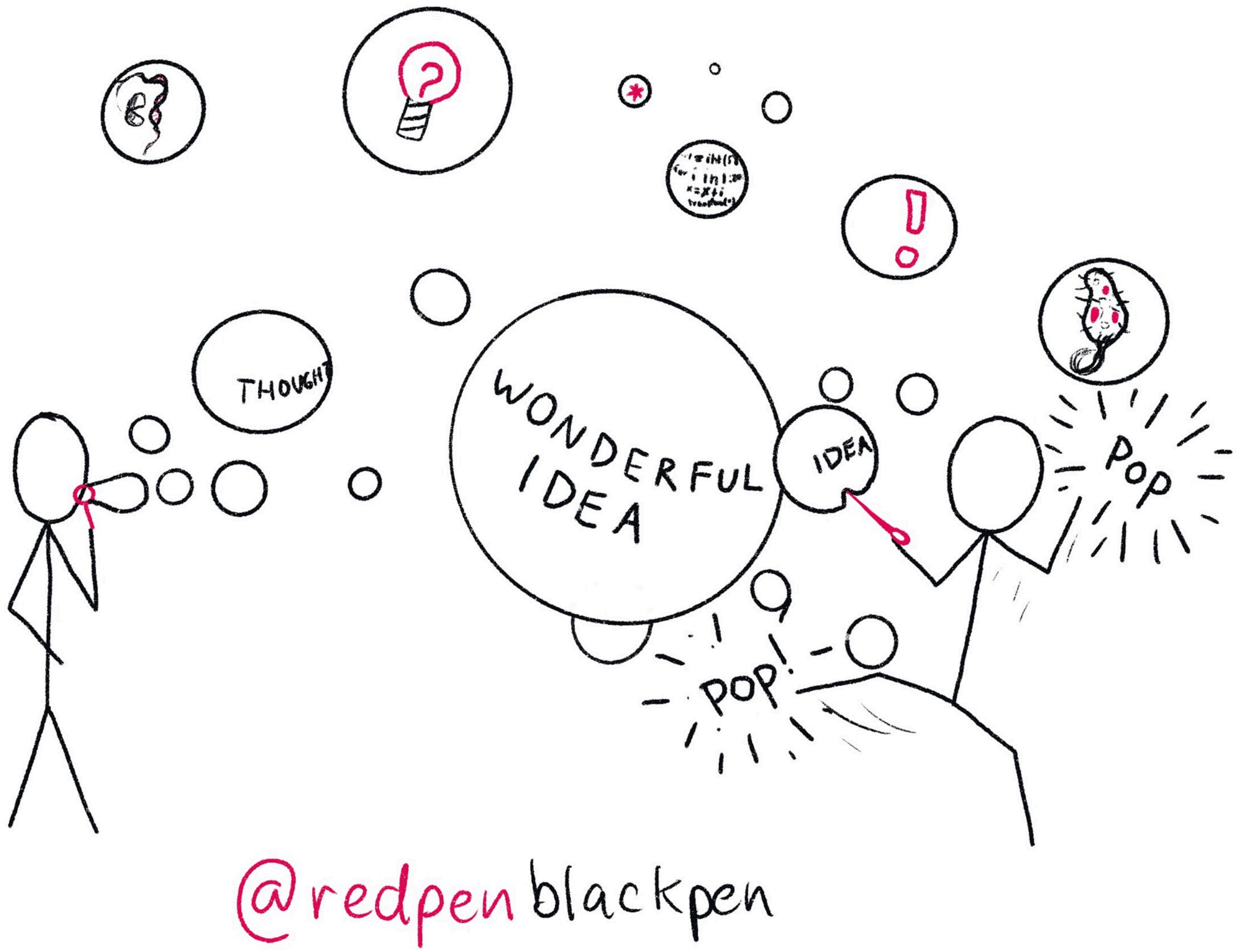



**Fig 1. Have ideas.** Ideas in the scientific space take a while to nucleate, but are ephemeral, like bubbles. When fully formed, good ideas will survive scrutiny and questioning. But be careful about exposing them to others too early. Criticism, however well-meaning, can burst half formed ideas.

https://doi.org/10.1371/journal.pcbi.1007448.g001

whatever that is, is very important (Fig 2). At some point, these scribbles do turn into ideas, but there is no magic wand. One analogy is of a nucleation point: initially there is a swirling mass of ideas with no form, and eventually they coalesce into something. Caffeine helps.

Discussing the idea with other people is vital, not just to shape the idea but also to find the right home for it. However, the timing of discussion is critical. Ideas are really fragile: Feedback at early stages will often focus on negatives, not the potential, and many a good idea has ended up in the bin due to early "help." Stephen King advises developing the initial idea with the door closed and only opening the door when the idea is mostly formed [2].

## Rule 2: Publish papers

Have no illusion: The main thing you need on your curriculum vitae (CV) is papers (Fig 3), preferably first-author papers and ideally first-author papers in which you are the corresponding author, with the occasional last author paper thrown in for good measure. Papers are both the imprint we leave on the scientific world and the genealogy by which other people can track our pedigree. A recent analysis identified papers as the single most important factor in getting tenure [3].

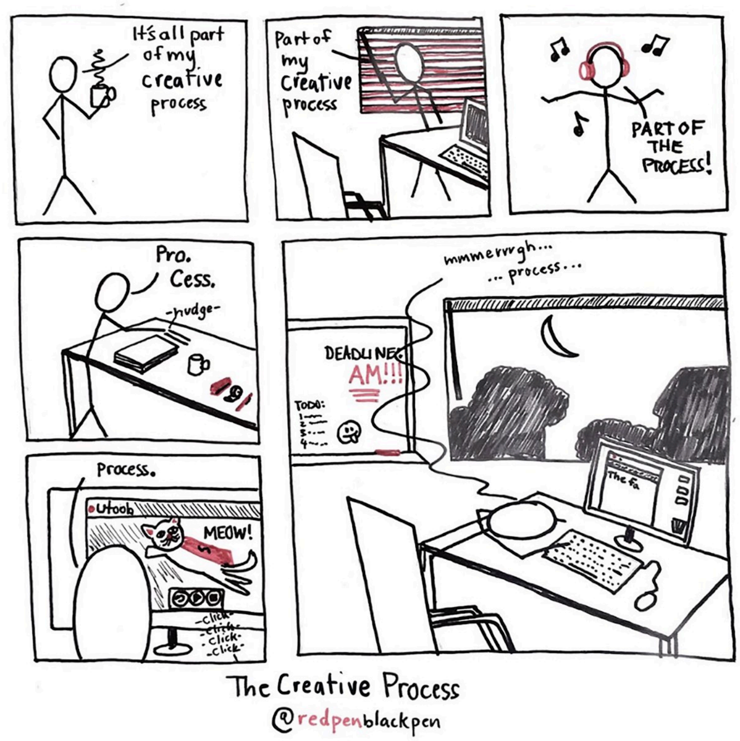



**Fig 2. The creativity process.** Writing is a part of the job, and it is important to have a process that helps you write (and good time-management skills around that process as well).

https://doi.org/10.1371/journal.pcbi.1007448.g002

While first-author papers are the gold standard, you also need breadth and depth in your publications. Look for opportunities to contribute to other people's work. Coauthored papers are important; they demonstrate an ability to collaborate as part of a team and can lead to roles in larger grants. One approach is to take advantage of what Dr. Tregoning calls "top drawer specials." Nearly every PI will have projects that never quite make it, due to people leaving or research priorities changing. Often all that is required is for someone to pull preexisting data into paper form. Completing these side projects whilst waiting for your main project to bear fruit is a very efficient way of CV boosting and practicing your writing.

You also need some evidence that you can get grants. Unfortunately, getting grants can feel like a catch-22 for early career researchers. You cannot become a PI till you get grants; you cannot get grants until you become a PI. This is a tricky but not insurmountable problem. There are some smaller pots of money that you can apply for, including travel grants and (sometimes) internal funding schemes. At the very least, be involved in grant writing. Learn the process, so it doesn't come as a horrible shock. If you do contribute to a grant from your current lab, ensure that you are named on it in some role.

Whilst there is no way that you can get an academic post without papers, papers alone are not sufficient: There are many people with great CVs and no tenure. There are other skills and behaviors that you need.

## Rule 3: Research what the job involves and learn to juggle

Before losing sleep about not becoming an academic, understand what an academic career involves. Spoiler alert: It is mostly juggling (Fig 4). Before becoming a PI, Dr. Tregoning drew

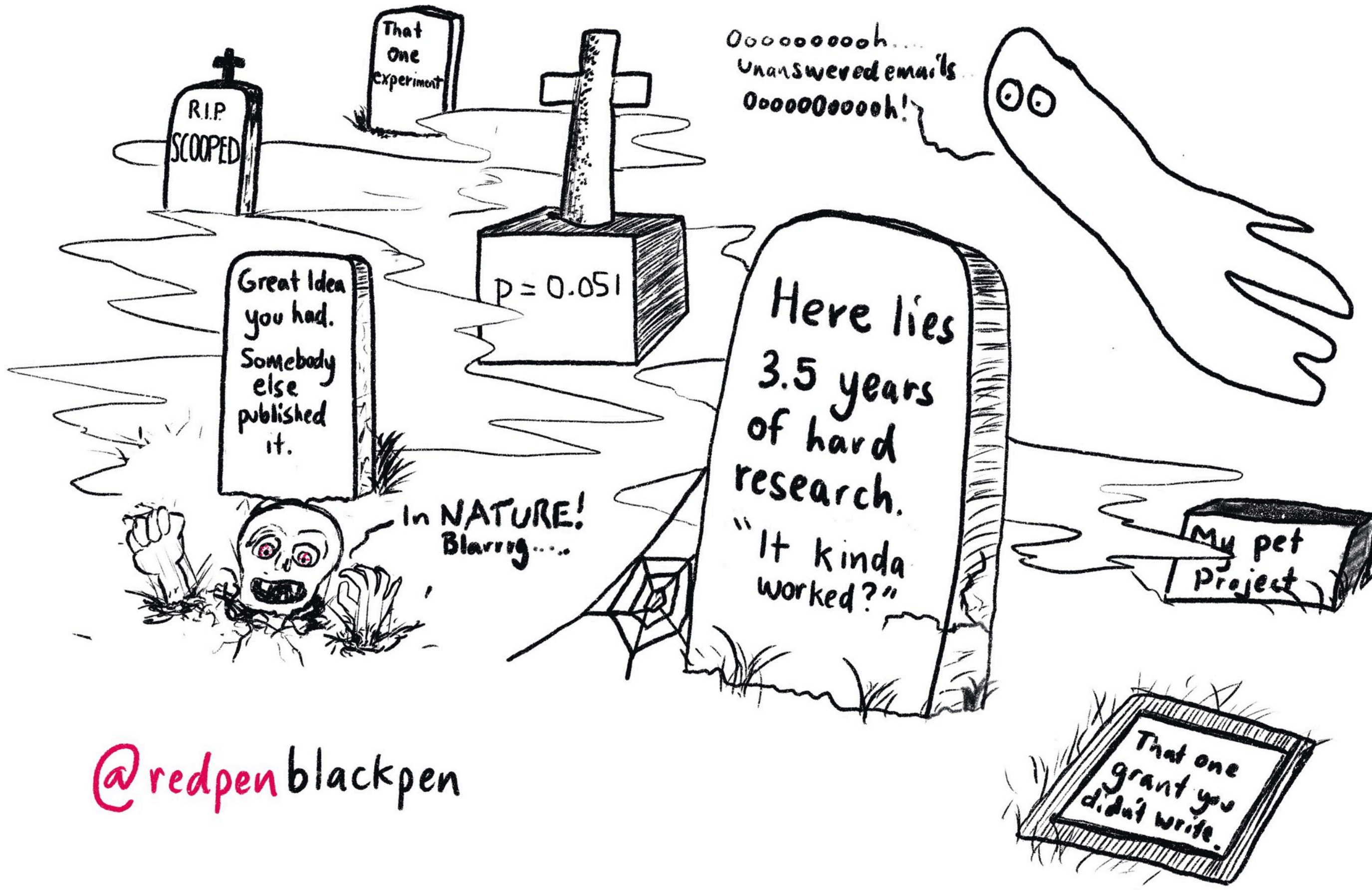



**Fig 3. Publish papers.** Every successful PI has a "graveyard" of uncompleted and/or unsuccessful ideas. The trick is to have a lot of things going forward to make sure some survive.

https://doi.org/10.1371/journal.pcbi.1007448.g003

heavily from fiction to form vague and entirely wrong ideas of the role, with elements of Hogwarts, Jordan College in *Lyra's Oxford* (from *His Dark Materials*), and the *Jurassic Park* cloning lab (before the dinosaurs escaped). However, a closer parallel is that you are an entrepreneur running your own business within an organization that provides some core support services.

As the head of a lab, you are responsible for fundraising, fund managing, purchasing materials and equipment (some of which is extremely specialist, even unique), training and managing staff working with dangerous materials, publicizing the current research, and planning future research. As a teacher you are expected to inspire and educate the next generation with a range of teaching styles that are appropriate for either 300 students in a lecture or for a single student, as a mentor. On top of this, you are expected to help with the administration of a large complex organization with upwards of 10,000 staff. Hiring good people can help to distribute some of this load. Dr. Hope Jahren in *Lab Girl* captures much of the joy and pain of an academic career [4].

No two academic careers are the same. This is one of the best aspects of the job. We are given (some) flexibility to choose our own routes. Whilst there are core elements—teaching, research, and administration—the make-up of each person's role can be very different. This will vary by individual and institution: Some places are research only, and some are focused on teaching. Spend some time thinking about what type of academic you might want to be and

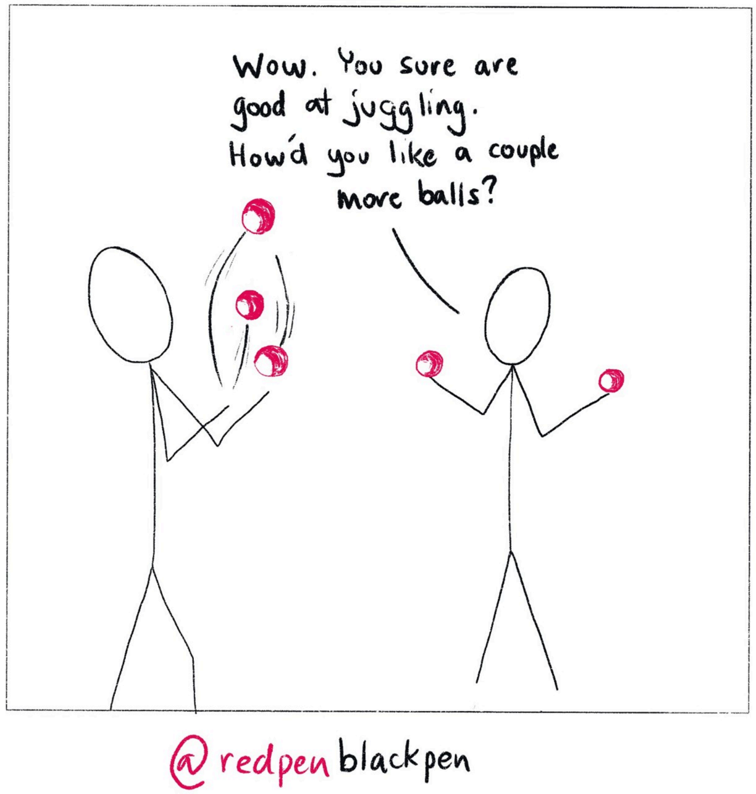



**Fig 4. Juggling, always juggling.** As a lab head, your duties may stretch your abilities to accomplish things. Be aware that if you are viewed as successful, you will be asked to take on even more.

https://doi.org/10.1371/journal.pcbi.1007448.g004

where your strengths lie so you can best prepare. But also try everything out: You might discover a previously unknown talent for steering committees, designing curricula, or community outreach.

## Rule 4: Develop your skills

The skills you learn in the lab during your PhD and postdoc are by and large irrelevant to those necessary to run a lab. While you will get your next job based on your CV (your previous employer and your publications), you will only succeed in your next job based on your ability to do a range of other things [5]. The biggest difference is lab and technical work. As a PI, the amount of time you spend doing raw science (be it in a wet lab, in a field, or at a computer) dramatically decreases. This can be tricky to come to terms with, but as the leader of the group, your main responsibility is to support your team. Invest some time in developing skills outside the lab. To get more of a sense of the skills needed to run a lab, read *At the Helm* [6].

The most important skill is learning to write well. The time that you no longer spend generating data is quickly filled by time writing grants and papers. Writing science well is not trivial. There are many resources that can support you in learning to write, including Stephen King's *On Writing* [2], Roy Peter Clark's *Writing Tools* [7], and Joshua Schimel's *Writing Science* [8]. There are also academic articles—including some excellent 10 Simple Rules [9]. If you do not have time to read these, take George Orwell's advice from "Politics and the English Language" and never say anything that is outright barbarous [10]. Nothing beats practice and feedback. Bear in mind, there are other ways to present your ideas [11].

The other critical skill is learning how to work with people. Get management experience before you go live with your own lab: That way, your early mistakes don't affect you long term. The easiest way is to do this is outside science, which can come in many forms—working in a shop, volunteering at a shelter, running a children's football team, or even joining the army [12], which may seem a bit extreme, but it gives you a chance to explore what works and what doesn't. Working with students is another rewarding way of developing your management skills. Likewise, ask your postdoc mentor if you can take on management responsibilities in your current lab.

## Rule 5: Focus on the prize

A lot of becoming a PI boils down to attitude: A major defining quality is relentless perseverance in the face of the odds. Your initial plan about how you're going to get there is usually a lot simpler and easier than the course you will eventually take, but don't give up (Fig 5). Whilst

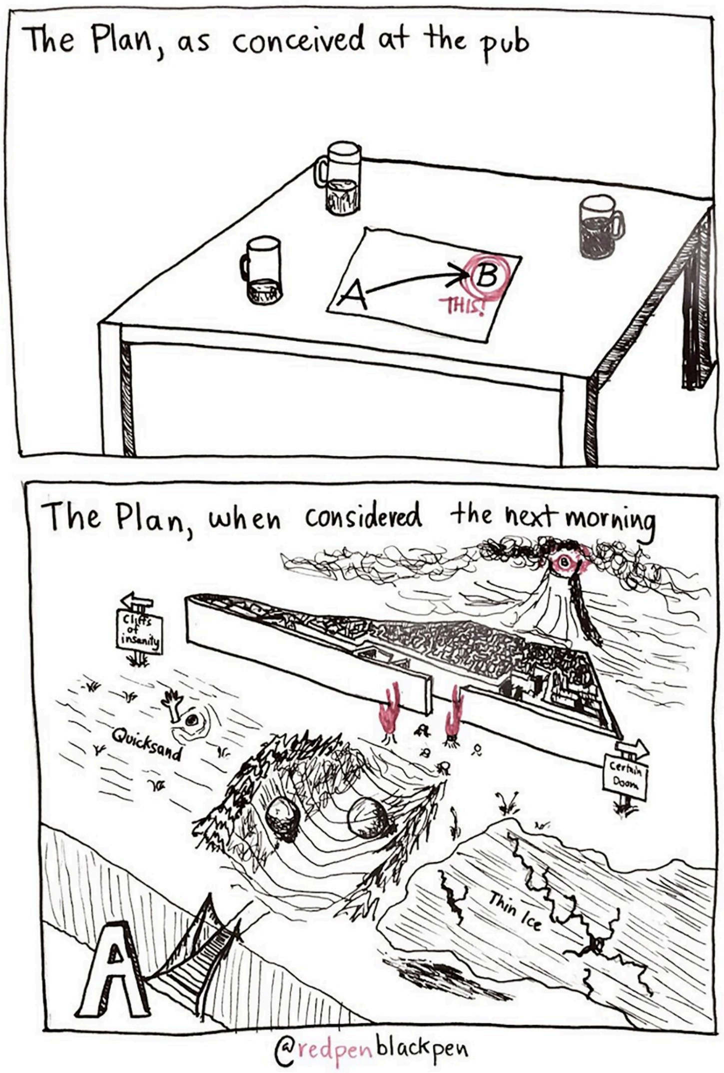



**Fig 5. Focus on the prize.** Generally, your plan seems a lot more straightforward than the way things actually happen. It's important to remain flexible about how you achieve your goals and, indeed, what those goals actually are. But also remember that other people's paths were not necessarily as smooth as they appear.

https://doi.org/10.1371/journal.pcbi.1007448.g005

you have limited hours in the day and, unfortunately, a limited time from getting your PhD to getting onto the tenure track, the solution isn't only working harder. Focus on the things that help you cross the line. In order to do this, you need to identify what these key things are and be able to evaluate the benefit per unit of time invested.

Make the most of your time: Think about what you are being asked to do and how it helps you become a PI. For example, be selective about the work you do: Don't start any work unless you can see the route to publication. One of the major skills is learning to say no, even if the opportunity is really shiny (Fig 6). This can be tricky when working for someone else who has ten scatty ideas that go nowhere for every one that makes the final draft. You have to be choosy and occasionally say no if the project looks like a dead end. Of course, there is a fine line between being self-driven and self-centred; you still need to do things that contribute to the smooth running of the lab you find yourself in. This extends into faculty positions. Being collegiate makes you more employable: No one wants to work with "that person" with a reputation for selfishness.

Be aware that what others show to the outside world isn't the whole truth. Professor Alice Prince at Columbia recently described how her National Institutes of Health (NIH) biosketch was a poor reflection of the person she is [13]. Going from a successful PhD studentship in a big lab to a postdoc in an even bigger lab, followed by a fellowship, with *Cell*, *Nature*, and *Science* (CNS) papers every step of the way is still perceived as the only route to being a PI. But very few people take this route; doing a muddle of short-term contracts is a much more realistic route. Trust us: It is possible to get a PI job without publishing in *Nature* (the authors submit their Google Scholar profiles as evidence: JST and JEM).

## Rule 6: Bounce back from failure

No matter how focused you are, you are going to fail. One of the most common experiences of being an academic is failure. You will fail on your path to becoming a PI, and you will fail once you become a PI. It is not the failing that matters; it is how you bounce back again (Fig 7). No one succeeds all the time; to use a sporting analogy, Babe Ruth had a batting average of 0.342—meaning he missed the ball 65% of the time—Lionel Messi requires 5.79 shots per goal, and Serena Williams misses 40% of her first serves. Likewise, a PI who gets more than 20% of their grants funded is a superstar. Learning coping strategies is vital. Some of the things that help are as follows:

### Reflective practice

Carol Dweck's brilliant book, *Mindset* [14], can help you to fail better. She suggests reframing failure as a learning opportunity. After the initial mourning period, look again at rejected papers and grants in the light of the reviewer's feedback and see what you can improve.

### Revise, recycle, and resubmit

Any single rejection is not the end of the idea. Many grant applications require a resubmission. In 2015, NIH R01-grant success rate was at approximately 30% for resubmissions, compared with 10% for original submissions [15]. And even if your grants are not funded, there are ways to get the work done, chopping them into new projects or running them on the side of other funded things. Likewise, no single journal is perfect, and with the San Francisco Declaration on Research Assessment (DORA) advocating for a shift from impact factor, it is worth considering on what other grounds your work will be evaluated [16].

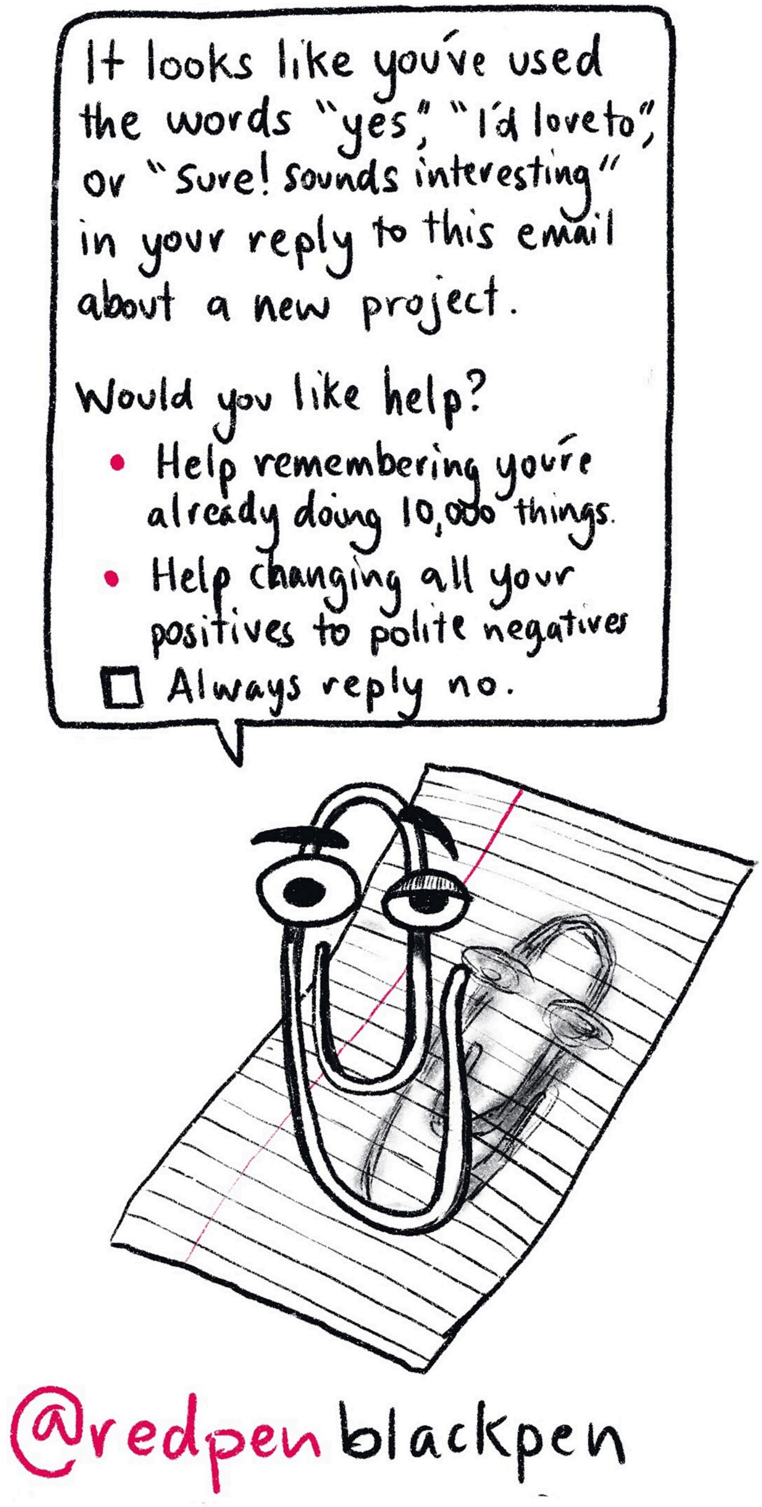



**Fig 6. Just say no.** The ability to say "no"—even when the opportunity seems exciting—is a must for PIs.

https://doi.org/10.1371/journal.pcbi.1007448.g006

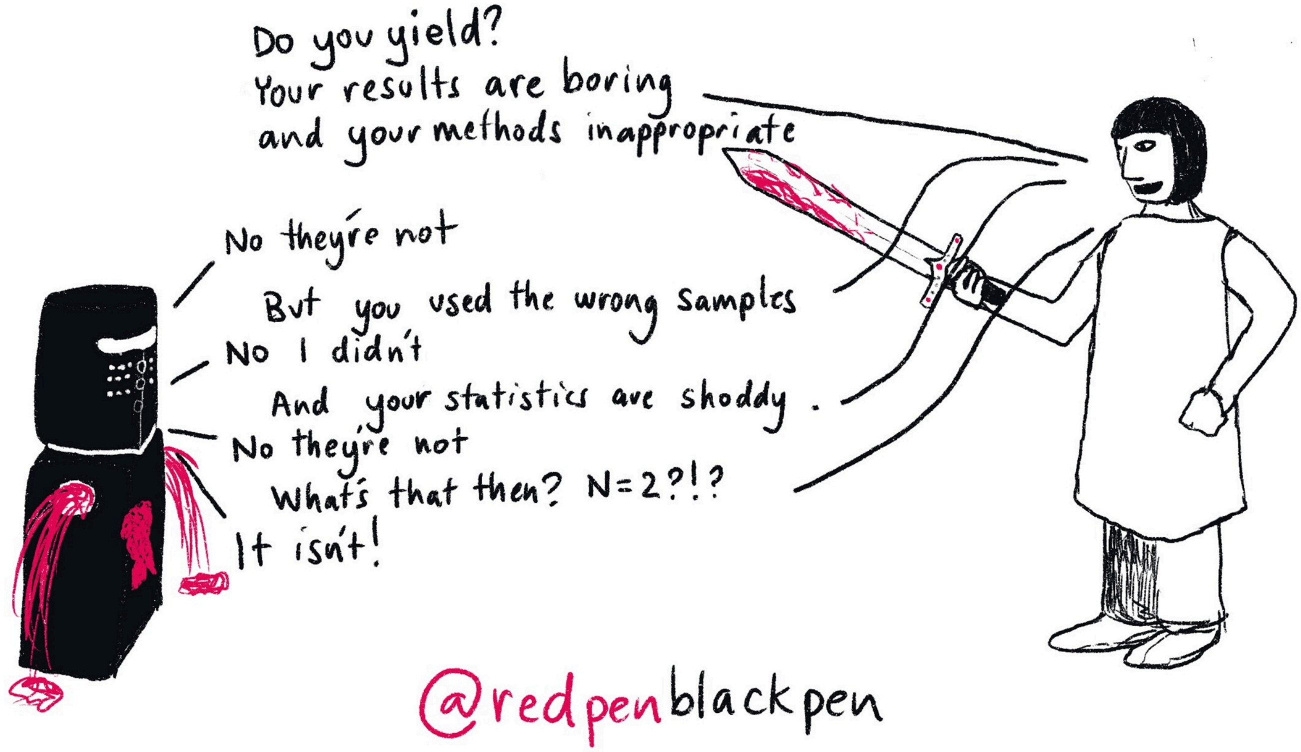



**Fig 7. Bounce back from failure.** Persistence and perseverance are two qualities that are vital to becoming and being a PI. You should stand up for and argue your point but remain aware that you could be wrong.

https://doi.org/10.1371/journal.pcbi.1007448.g007

### Learn to roll with reviewers' punches

Flawed as it is, the peer-review system is the least worst, and we are stuck with it for now. Remember, reviewers don't reject your work because they hate you. Stand in the reviewer's shoes: They have to make decisions on a large number of grants from a multitude of subjects in limited time. Sometimes, your work may not make the cutoff, and coming back a second time with new data may be enough to get you over the line.

### Depersonalize it

It is important to separate your personal worth from your successes and failures at work. Failure and the resultant iterations and corrections are part of creativity. Depersonalizing failure allows you to accept constructive criticisms and move both your ideas and yourself forward [17].

## Rule 7: Develop your brand

We have two things to sell, our ideas and ourselves. Of the two, the main product we sell is ourselves, which (at work at least) is defined by our CV: what we have worked on and who we have done it with and where. Develop a single memorable "personal brand," which can be used when meeting potential collaborators, conference organizers, and funders. Have a single line "elevator pitch" that summarizes what you do, backed up with an exciting case study. The brand includes the types of research you aspire to do and the initial projects you might run. The hope is that by pitching this brand successfully you will be at the forefront of people's thoughts when they are putting together grants, consortia, or seminars. Your brand could even help you end up in front of that elusive tenure-track or lectureship appointment committee. Part of this brand development is identifying your strengths and honing them. Whilst you shouldn't ignore your weaknesses, your strengths are the foundation on which you build your career. The brand is no longer limited to papers and conferences. It is possible to reach whole new audiences through social media [18]—though be aware the boost in connectivity may not compensate for the time lost down the rabbit hole.

## Rule 8: Believe in yourself

Developing your brand is easier said than done, in part because of the curse of imposter syndrome [19], in which you doubt your own talents and fear that you will be revealed as a fraud (Fig 8). Nearly everyone in academia suffers from it to some degree or other. The process of peer review is a major contributor: You and your work are routinely judged by others and, given the high failure rate, are going to be found wanting, often.

At the end of the day, academia is just a job, a fascinating and fun job that can occasionally take up every single second you have but ultimately just a job. There is no value in becoming a PI at any cost. Ensure you live a life outside work, for your own mental health and for that of your team. Academic burnout is very real; it is okay to take a break to reset stress levels: for example, Dr. Tregoning runs [20]. Sometimes stepping back a bit can even help in terms of creativity and headspace. Likewise, if you work every hour of the day to become a PI, there is no spare capacity when life inevitably doesn't run smoothly.

## Rule 9: Build a network of mutually supportive people

Networking is central to being a PI. The best way to do this is to meet people face to face: Get out there and break bread. Carry business cards at all times. Go to conferences, consortia, and congresses: Plan who you want to meet at the conferences, even email them in advance to arrange time at the meeting. Often, smaller conferences can give you better chances to meet

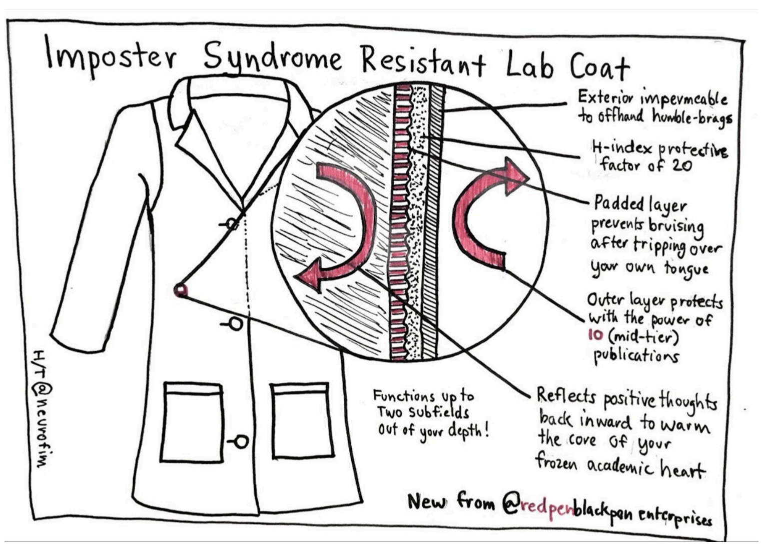



**Fig 8. Believe in yourself!** You can make one of these lab coats for yourself by trying to remember that everyone in science experiences imposter syndrome to some extent and that questioning yourself, your perspective, and your position is actually an essential part of the scientific process.

https://doi.org/10.1371/journal.pcbi.1007448.g008

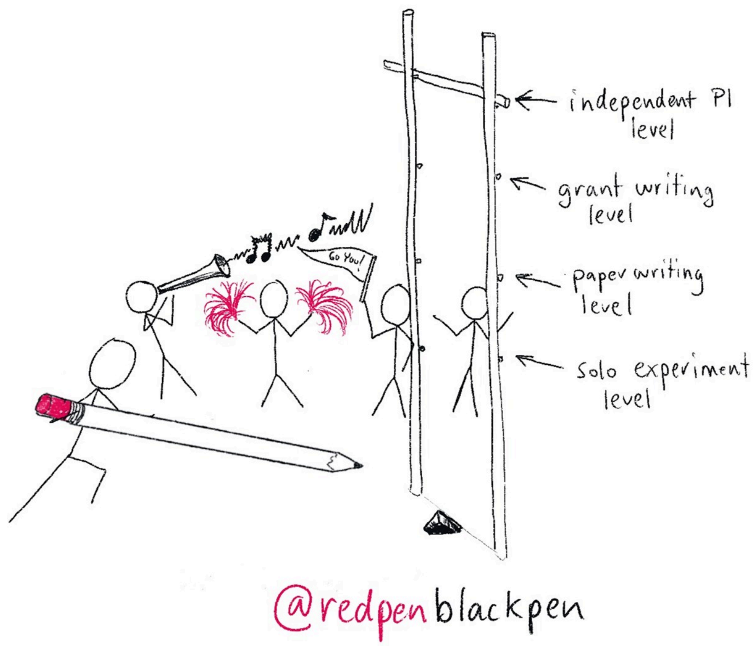



**Fig 9. Get help.** Remember that sometimes you need the cheering section and sometimes you are the cheering section for others. PI, principal investigator.

https://doi.org/10.1371/journal.pcbi.1007448.g009

people. But be aware that talks aren't the only part of the conference; the social events are great places to meet people and to learn. Networking isn't limited to networking up; network sidewise with your peers and down with the people who you are training. Virtual networking can help: This paper is the offspring of a Twitter conversation.

The other consideration is choosing the right boss and environment to work in. The ideal boss is supportive, enabling, and generous in credit. If you can't find that, find someone who will let you get on with things independently so you can develop your own ideas. At the very least, avoid bosses who will wittingly or unwittingly damage your career. Try to discover what flavor a potential boss might be before committing to work for them; discrete questions when you visit a lab for an interview can be helpful.

Ultimately, nobody can succeed on their own (Fig 9). There are many functional reasons to build up a network of people: Other people have different skills, expertise, and access to equipment or reagents. Many of the best things that happen are often random offshoots from chance meetings, for example, papers that sprang from discussions with the external examiner at a viva or collaborations formed at conference bars.

But, more importantly, network because working on your own is rubbish, boring, and sad. Nurture your colleagues at every step of the way. Be kind wherever possible [21]. You could probably succeed by pursuing a divide and conquer, winner takes all attitude, and you could probably name people who have done this. But don't; as then, everyone loses.

## Rule 10: Know when to quit

Disclaimer: Even if you follow these 10 rules, there is no guarantee you will end up running your own lab. Knowing when to cash out is as true for scientific dead-ends as career ones. Chasing after something long after it has eluded you is not going to add to your overall life happiness.

Remember that academia isn't the only job. There is the perception that becoming a PI is the one true path. As a PhD student, it is easy to say that your dream goal is to become a PI. This is mostly groupthink resulting from the lack of visible alternatives. Within the university system, the most visible individuals who are successful are the PIs; the people who are succeeding in other careers have by necessity left the university system. Becoming a PI is merely one career path amongst several that are available to scientifically trained graduates, all of which will value the skills you have developed along the way. Make sure that in upskilling you consider employability outside the sector or giving yourself an edge within the sector; this is part of the reason the authors started blogging and drawing: It was a new thread drawn from skills we enjoyed that gave us a different dimension [22].

Deciding when enough is enough is the hardest rule and is in direct conflict with Rule 6 about learning to fail. This is a decision only you can make, but don't rush it as it is not unusual to want to quit often.

## Conclusion

It has not been easy to condense how to become a PI into Ten "Simple" Rules, particularly when there are so many tensions—being focused on yourself without being selfish, being resilient in the face of failure but knowing when to quit, and gaming it without being cynical. Notable absentees from this list are technical expertise, hard work, and knowledge of the field: These are a given, but there are more people who are hardworking, skillful, and knowledgeable than there are PI jobs. There is clearly a role for luck, but you need to be prepared to exploit opportunities.

Ultimately, if becoming a PI is what you want to do, do not let anyone (including yourself) put you off. Yes, the odds are against you, with a recent study reasserting the low rate of postdocs who become tenure-track faculty [23], but there are jobs with worse, steeper pyramids. Tracking the likelihood of reaching the top in other careers is one way to normalize academia: for example, acting—of the 300 million people in the US only 51,000 people work as actors (Actors' Equity Association figures 2017). Likewise, while nearly every child in England plays soccer at some point in their childhood, only 22 of them make the national teams. Compared to these horrific odds, academia is relatively easy: 15% of the roughly 20,000 postdocs employed in the US will end up in a tenure-track academic position [24].

Ultimately, it is social intelligence (sometimes referred to as 'soft skills') that can make the critical difference. The good news is that you are already developing a lot of these skills by stealth: Time management, working with people, and juggling priorities are all part of being a postdoc. Even better, these leadership skills—being more resilient, being kind, looking after yourself and your colleagues, and focusing on your goals—apply to all jobs. So even if your academic aspirations don't play out, you will be in a position to succeed in any role.

## Acknowledgments

The authors are grateful to Emily Graham (Pacific Northwest National Laboratory) for the invaluable feedback.

## References

1. Savage A. Every tool's a hammer: life is what you make it, 2019.
2. King S. On writing: a memoir of the craft. Scribner trade paperback edition. ed. New York: Scribner; 2010. x, 291 pages p.

3. van Dijk D, Manor O, Carey LB. Publication metrics and success on the academic job market. Current Biology. 2014; 24(11):R516–R7. https://doi.org/10.1016/j.cub.2014.04.039 PMID: 24892909
4. Jahren H. Lab girl. First edition. ed. New York: Alfred A. Knopf; 2016. 290 pages p.
5. Ranganathan S. Towards a career in bioinformatics. BMC Bioinformatics. 2009; 10(15):S1. https://doi.org/10.1186/1471-2105-10-S15-S1 PMID: 19958508
6. Barker K. At the helm: leading your laboratory. 2nd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2010. xii, 372 p. p.
7. Clark RP. Writing tools: 50 essential strategies for every writer. College ed. Washington, D.C.: CQ Press; 2008. xv, 260 p. p.
8. Schimel J. Writing science: how to write papers that get cited and proposals that get funded. Oxford; New York: Oxford University Press; 2012. xiv, 221 p. p.
9. Dashnow H, Lonsdale A, Bourne PE. Ten Simple Rules for Writing a PLOS Ten Simple Rules Article. PLoS Computat Biol. 2014; 10(10):e1003858. https://doi.org/10.1371/journal.pcbi.1003858 PMID: 25340653
10. Orwell G. Politics and the English Language. The Broadview Anthology of Expository Prose. Toronto: Broadview Press; 1946.
11. McDermott JE, Partridge M, Bromberg Y. Ten simple rules for drawing scientific comics. PLoS Comput Biol. 2018; 14(1):e1005845. https://doi.org/10.1371/journal.pcbi.1005845 PMID: 29300733
12. Tregoning J. From parade ground to PI. Science. 2018; 359(6373):362. Epub 2018/01/20. https://doi.org/10.1126/science.359.6373.362 PMID: 29348239.
13. Prince A. Omissions from a National Institute of Health (NIH) biosketch. PLoS Pathog. 2018; 14(5): e1006896. Epub 2018/05/11. https://doi.org/10.1371/journal.ppat.1006896 PMID: 29746562; PubMed Central PMCID: PMC5944916.
14. Dweck CS. Mindset: the new psychology of success. Updated edition. ed. New York: Random House; 2016. xi, 301 pages p.
15. Lauer M. Resubmissions Revisited: Funded Resubmission Applications and Their Initial Peer Review Scores 2017. Available from: https://nexus.od.nih.gov/all/2017/02/17/resubmissions-revisited-funded-resubmission-applications-and-their-initial-peer-review-scores/. [cited 2020 Jan 17]
16. Tregoning J. How will you judge me if not by impact factor? Nature. 2018; 558(7710):345. Epub 2018/06/21. https://doi.org/10.1038/d41586-018-05467-5 PMID: 29921857.
17. Su F. The Lesson of Grace in Teaching. From The Mathematical Yawp blog 2013 [Internet]. Available from: http://mathyawp.blogspot.com/2013/01/the-lesson-of-grace-in-teaching.html. [cited 2020 Jan 17].
18. Britton B, Jackson C, Wade J. The reward and risk of social media for academics. Nature Reviews Chemistry. 2019. https://doi.org/10.1038/s41570-019-0121-3
19. Persky AM. Intellectual Self-doubt and How to Get Out of It. Am J Pharm Educ. 2018; 82(2):6990–. https://doi.org/10.5688/ajpe6990 PMID: 29606718.
20. Tregoning JS. Running up that hill. From Dr John Tregoning, Academic blog [Internet]. Available from: https://drtregoning.blogspot.com/2018/05/running-up-that-hill-running-pop-songs.html [cited 2020 Jan 17].
21. Tregoning J. No researcher is too junior to fix science. Nature. 2017; 545(7652):7. Epub 2017/05/05. https://doi.org/10.1038/545007a PMID: 28470218.
22. McDermott J. Drawing connections. Science. 2017; 356(6343):1202. Epub 2017/06/18. https://doi.org/10.1126/science.356.6343.1202 PMID: 28619947.
23. Hayter CS, Parker MA. Factors that influence the transition of university postdocs to non-academic scientific careers: An exploratory study. Research Policy. 2018. https://doi.org/10.1016/j.respol.2018.09.009.
24. Puljak L, Sharif WD. Postdocs' perceptions of work environment and career prospects at a US academic institution. Research Evaluation. 2009; 18(5):411–5. https://doi.org/10.3152/095820209x483064

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules to aid in achieving a vision

**Philip E. Bourne**[1,2]*

1 School of Data Science, University of Virginia, Charlottesville, Virginia, United States of America, 2 Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America

* peb6a@virginia.edu

> *Whoever has visions should go to the doctor.*
>
> Helmut Schmidt [1]

In a career that now spans 40 years, I have had, on several occasions, opportunities to turn loose ideas into a unified vision and the resources to implement that vision. What do I mean by a vision? A vision, at least in my mind, is the ability to see something important to the future, perhaps before others do. Fulfilling that vision does not have to change the whole world (although that would be nice) but only to impact others in a positive way. Consider what I perceive has been my own visioning to provide some context.

My visioning began in the 1990s when, seeing what computation was bringing to the life sciences through work on the human genome project, I was lucky enough to be able to envision and establish a bioinformatics laboratory before the idea became mainstream. In the early 2000s, it was a collective vision for what an exemplar data resource, namely, the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB), should achieve. Around 2005, it was something dear to this readership, a vision for a new journal, *PLOS Computational Biology*, for which I was cofounder and Founding Editor-in-Chief for 7 years. Around 2007, it was forming a company, SciVee.tv, to envision how digital media other than print could be used to communicate science. Around 2014, as the first Associate Director for Data Science (ADDS) for the United States National Institutes of Health (NIH), the vision was how big data could catalyze change in life sciences research. Finally, now in 2019, the vision is how one of the first academic Schools of Data Science should be established and run. These either were, or are, great opportunities to lay out a vision and act upon that blueprint. Not all were successful (*PLOS Computational Biology* was), but all were learnt from. So, what did I learn? Here is at least part of my life's lesson, in that now familiar and comfortable Ten Simple Rules format. In this article, the rules are generic and can be considered beyond our own discipline.

## Rule 1: Be realistic

I went to the US NIH with a vision to create a truly interdisciplinary environment, breaking down the silos (i.e., cylinders of excellence) that exist in individual institutes using data sharing as a catalyst. This was an unrealistic vision, given that such a change was not about better technologies; it was about a change in culture, and cultural changes typically occur slowly and in increments. Fulfilling a vision requires being realistic as to what can be accomplished and in what timeframe, yet it also means seeing beyond the current culture. It is all too easy to accept the status quo, and doing so will not allow you to envision anything but simply be incremental in your thinking. So, it is necessary, in the case of the NIH and indeed any academic institution, to first continually and consistently communicate your vision (see Rule 6) and figure out

Check for updates

OPEN ACCESS

**Citation:** Bourne PE (2019) Ten simple rules to aid in achieving a vision. PLoS Comput Biol 15(10): e1007395. https://doi.org/10.1371/journal.pcbi.1007395

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** October 17, 2019



**Funding:** There was no source of funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

ways to slowly introduce components of the overall vision that are easily seen to make sense to the organization writ large.

## Rule 2: Truly believe in your vision

If you don't wake up thinking of how to fulfill your vision, you are likely not going to. That waking thought is a reflection of your passion and belief. The passion you display will be associated with many of the other rules that follow. If what you propose truly is visionary, there will be many who will be the first to tell you why it is a bad idea. If your belief falters based on their input, you are not going to achieve the vision. Your own convictions begat the strength and courage to move forward against the odds. To quote one example from my own experience, when we took over the RCSB PDB from the Brookhaven National Laboratory, there were naysayers both verbally and in print who said this was "the end of structural biology." There was great scrutiny and community pressure to succeed. If we had not truly believed in our collective vision (see Rule 7), we could have faltered.

## Rule 3: Have a plan to fulfill the vision and expect both to change

A vision must be more than a vague idea. It will start that way, but if you can't very quickly write down a plan for how to fulfill that vision, it likely is not going to amount to much. You certainly won't have it all worked out at the outset, but if you can't build on the basic concept as you begin to popularize and evangelize your vision (see Rule 6), then it is likely doomed. As you do start to flesh out the vision, the passion (as defined by Rule 2) should increase. If it diminishes, you are in trouble. The vision will also morph as others weigh in and get involved. That is usually a good thing as long as that influence enables you to see the vision more and not less clearly. Another outcome is how time morphs your vision. Case in point was this journal. The original vision was to get experimentalists to read and adopt the work of computational biologists. I am not convinced that happened to the degree hoped; what did happen was that, over time, we saw an ever-increasing number of scientists trained in computation and experiment. The end result was the same: success—a highly sought-after journal, with lots of other innovations that were not part of the original vision (this Ten Simple Rules series being an unforeseen example).

## Rule 4: Be in the right place at the right time and seize the opportunity

I have pitched a vision for a company in the hope of getting funding on the very same day that another company doing much the same publicly announced, with a big splash, their first round of funding. We were much too late to the marketplace. Conversely, being too early will likely not enable others to see the vision—there is just not enough context for others to grasp, regardless of how hard you work on Rule 6. However, don't forget those "too early" visions completely; their time may come. I wish I had advice for knowing what is the right place and time, but I do not. It's mostly luck. In a crude way, it relates to Rule 5. If you can't fund the vision, then it is not the right place or time.

## Rule 5: Be able to finance your vision

A vision will likely never be realized without money. Without money, it will be a hobby at best. This is certainly true of a scientific vision. Money can come from grants, contracts, angel or venture capital investment, philanthropy, and I am sure other sources I have yet to imagine. A vision is a big idea, and it will cost. Rule 6 comes into play here. Others have to see the value in your vision, whether they be grant reviewers, big time venture capitalists, philanthropists, or others. The vision needs to align with the requirements of the funding source. Anyone who has written and then not gotten a grant knows this. Less obvious is alignment with other funding sources. For example, when we tried to raise money from venture capitalists with what we thought was a great vision for a new form of scientific communication, they did not agree, asking "where is the hockey stick?" implying they did not see the financial growth potential after an initial investment. We showed only a steady growth, not the next Google or Facebook. On the other hand, we just received the largest gift in our university's history as our vision for a School of Data Science was in alignment with that of the donor foundation. That alignment often comes from trust built over time.

## Rule 6: Be able to communicate the vision effectively

Chances are, a meaningful vision will be outside of the scope of what those who can carry it forward can easily comprehend, otherwise they would have thought of and possibly implemented that vision themselves. A clear articulation of what your vision is, how it will be implemented, and most importantly, what positive effect it will have are critical. Again, drawing on my own experience, our articulation of the NIH Data Commons [2], which abided by the Findable, Accessible, Interoperable, and Reusable (FAIR) data principles [3], illustrates this rule. It also illustrated the importance of communication to all stakeholders—from institutional leadership in supporting the vision to those who are the beneficiaries of its implementation. As an aside, the vision for the NIH Data Commons also brings up two other points, the first related to Rule 1. The NIH Data Commons was only part of what we envisioned for the NIH and came from being realistic. Second, the vision does not necessarily need to be a completely new idea—the Genomic Data Commons [4], for example, already existed.

Articulation of the vision comes in multiple forms—written and verbal and in multiple media, formal peer-reviewed articles, podcasts, social media, slides and associated presentations, and video. Once you have articulated the vision, keep doing so. If the vision makes sense, sooner or later, others will start to articulate the same vision as if it were their own or, indeed, articulate an even better vision, which takes us to Rule 7.

## Rule 7: Be willing to share ownership

Although you might have the original vision, you will not carry it out alone. Sharing the vision also involves being willing to share the credit for that vision or, indeed, to gain no credit for that vision at all. If you are not willing accept these outcomes, chances are your vision will die on the vine. Sharing with others who have their own visioning skills will amplify the value of what you foresee. Once shared, suitable partners will also communicate the vision, and so the value of the vision compounds while your contribution as an individual diminishes. Sharing is antithetical to the academic system that is geared (alas) towards rewarding the individual, and so traditional reward systems may not apply. For the visions I have been involved in, they just would not have happened without sharing and collaboration. Consider my latest attempt at a vision: envisioning a new type of school in a conservative academic setting. If successful, this will be the result of senior leaders, donors, and a team that shares the original vision, shares their additive contributions to the vision, and pulls together as a team.

## Rule 8: Be patient and don't expect rewards

The greatest of visions takes time to be adopted. A vision is typically not a single discovery, although such a discovery might be foundational to that vision. The structure of DNA was foundational, the discovery of CRISPR-Cas9 was foundational, but the vision lies in how those discoveries can be broadly used, which takes time to be appreciated (in these examples, genomics more broadly and the possibilities of gene editing, respectively). Science typically recognizes discoveries but not necessarily the vision to utilize those discoveries. The reason, as already stated, is that science typically rewards individuals but not a collective vision. Stated another way, discovery belongs to a few individuals; visions that build upon those discoveries are many and belong to many people.

## Rule 9: Know when to stop

You may have a great vision, but it becomes clear that it is just not going to happen. The reasons can be many. One I have experienced a number of times and raised here in Rule 1 and elsewhere relates to a vision that requires a change in culture. Changing culture in anything less than a generation is hard. It can happen, but the incentives have to be there, or you have to be able to find ways to change the incentives, which also can be very hard. Consider interdisciplinary research, which cuts across a number of my own visions. Most would agree that there is much to be gained from interdisciplinary research, but still, outdated models of funding and university organization hamper interdisciplinary activity. Money still flows to the traditional silos; any vision that seeks to break that model and the culture it implies will be hard pressed. That does not imply that one should give up on the vision; just be more realistic as to the time it will take. Knowing which visions are a lost cause as they will never happen and those that will take a long time is something I am still learning. Looking back, I would say that when the pace of seeing your vision come to fruition is so slow as to have the frustration it causes impact the rest of your life, it is time to stop and take up a new vision.

## Rule 10: Take time to enjoy the experience

Fulfilling a vision requires immersion in the process of making that vision a reality. This can become all-encompassing to the point that when one milestone is reached, rather than taking the time to relish that success one rushes on to the next challenge. It's easy to become obsessive and not enjoy the creative experience. In accordance with Rule 7, you will have others with whom to share the experience. Take the time to do so. It builds the team and makes it all worthwhile. In our latest vision for establishing a School of Data Science, we celebrate each milestone together with those who have made it happen. This takes the form of dinners, sometime at my house, which speaks to a commitment of time and the notion of family. In the case of *PLOS Computational Biology*, it is the annual dinner for editors as an annual meeting. In the case of the lab, it is hikes, fishing, skiing, and other social occasions to celebrate together and with our families.

We all envision; we would not be scientists if we did not. Visions come at different scales and in different forms. Certainly, these rules are neither sufficient nor will they cause you to fulfill your visions. Hopefully, they will at least get you thinking about the visioning process, which is different for each one of us. To that end, the rules are generic and broadly applicable across science and, indeed, society.

As a scientist, all I can do is offer my own experiences and suggest that, although the body of work in your next paper is a vision of sorts, you should look beyond that to a body of work, the translation of that body of research into product, and the influencing of your whole field. As a society, our future depends on it.

## Acknowledgments

Thanks to Daniel S. Katz for making this article better and proving that collaboration is important, not just for carrying a vision forward but also for writing about visions.

## References

1. Helmut Schmidt Wikipedia page. Available from: https://en.wikiquote.org/wiki/Helmut_Schmidt. Last updated 2018 September 24. [cited 2019 September 28].
2. New Models of Data Stewardship. NIH website. US Department of Health and Human Services. https://commonfund.nih.gov/commons. Last updated 2019 May 13. [cited 2019 September 28].
3. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016; 3: 160018. https://doi.org/10.1038/sdata.2016.18 PMID: 26978244.
4. The Next Generation Cancer Knowledge Network. National Cancer Institute Genomic Data Commons website. US Department of Health and Human Services. https://gdc.cancer.gov/. [cited 2019 September 28].

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for giving an effective academic job talk

**Shayna A. Sura[1‡*], Lauren L. Smith[1‡], Monique R. Ambrose[1‡], C. Eduardo Guerra Amorim[1‡], Annabel C. Beichman[1‡], Ana C. R. Gomez[1‡], Mark Juhn[1‡], Gaurav S. Kandlikar[1‡], Julie S. Miller[1‡], Jazlyn Mooney[2‡], Riley O. Mummah[1‡], Kirk E. Lohmueller[1], James O. Lloyd-Smith[1*]**

**1** Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, United States of America, **2** Department of Human Genetics, University of California Los Angeles, Los Angeles, California, United States of America

‡ SAS, LLS, and MRA contributed equally to this work. CEGA, ACB, ACRG, MJ, GSK, JSM, JM, and ROM also contributed equally to this work.
* ssura@ucla.edu (SAS); jlloydsmith@ucla.edu (JOLS)



**Citation:** Sura SA, Smith LL, Ambrose MR, Amorim CEG, Beichman AC, Gomez ACR, et al. (2019) Ten simple rules for giving an effective academic job talk. PLoS Comput Biol 15(7): e1007163. https://doi.org/10.1371/journal.pcbi.1007163

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** July 25, 2019



**Funding:** The authors acknowledge support from the National Science Foundation Graduate Research Fellowship Program (to SAS, ACB, and JSM under grant #DGE-1144087, and to GSK and JSM under grant #DGE-1650604), the NSF Postdoctoral Research Fellowship in Biology (to JSM under grant #DBI-1812292), and NSF research grants OCE-1335657 and DEB-1557022 (to JOL-S and ACRG). ACRG was supported by the CAPES Science Without Borders Doctoral Fellowship. ROM and JOL-S were supported by the US Department of Defense Strategic Environmental Research and Development Program (RC-2635). The funders had no role in study design, data

## Introduction

You've finally completed your dissertation research and have your PhD in hand—yay! Maybe you're also in the middle of a postdoctoral position. If you're reading this article, chances are you are actively searching for and applying for faculty positions. (Check out reference [1] if you're early in the application process and [2] for additional advice!) Unfortunately, many graduate students and postdocs are not taught the skills necessary for acquiring a faculty position after passing the "looks good on paper" part of the application and securing an on-campus interview. One of the last crucial steps in earning a faculty position is your academic job talk. No matter how great of a scientist you are, if you cannot give a compelling job talk, chances are low that you will be hired. Yet many candidates receive little guidance on how to ace this unique and vital test.

To help address this gap, we have put together these ten simple rules that will help you give an effective job talk. To be clear, these are rules developed for the academic job talk in a research-heavy department, which is typically in a seminar format. These rules are not targeted toward other formats such as chalk talks or teaching demonstrations, although some pointers may still apply. We are a group primarily composed of University of California, Los Angeles (UCLA) faculty, postdocs, and graduate students who participated in two recent job searches in the Ecology and Evolutionary Biology Department. We evaluated ten job talks over the span of 2 months and discussed their strengths and weaknesses in a weekly seminar course. These ten rules are based on our discussions of what worked (and what didn't) across the variety of job talks we observed, as well as our various experiences on the job market and search committees over the years.

## Rule 1: Know your audience

As with any seminar or presentation, when preparing your job talk, you want to target your specific audience. Therefore, you need to consider the background knowledge and interests of the audience members. Learn as much as you can about the position and what institutional needs the position is meant to address within the department and broader university. If you're applying for a position within a specific department, what is the scope of the research in that department? Does it have a mission statement? Are any strategic aims or future plans publicly available? Does the department work closely with other academic units on campus, and does the position you've applied for have any formal ties to other units? To answer some of these questions, you should read the job ad closely, read about the current faculty's research, and

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

look through the department's web page (see also Rule 7 [Understand your potential new workplace] and 8 [Understand your new colleagues] from reference [3]). If you're lucky enough to have network connections to the department, use them now to get insights before you visit. We also recommend that after you receive an invitation to interview, you consider setting up a phone call with the chair of the search committee to inquire about the job and ask any specific questions you have regarding the job or department. In particular, it is a good idea to ask what the search committee is looking for—it may have been a long time since the job ad was released, and the search committee's focus may have shifted from what was initially stated. We recommend a phone conversation as opposed to an emailed list of questions because it saves time; also, people are often more candid and may provide more useful insights over the phone. Depending on when your job talk occurs during your interview schedule, you might even make small changes to customize your talk based on interviews and meetings with department members prior to your talk.

## Rule 2: Sell yourself

The faculty and search committee are trying to choose the candidate they'll be most excited to have as a new colleague, so you need to showcase the reasons you're their best choice! It is smart to include an explicit introduction about yourself—i.e., the kind of science you do, your grand aims, and your approach to research. You want to communicate your identity as a researcher and, if appropriate given your career stage and research plans, how this differentiates you from your mentors (reference [4] is an excellent resource).

You also want to convey other traits as a scientist and potential colleague. Reflect on the qualities that make you an exceptional researcher (creative, persistent, thoughtful, rigorous, multidisciplinary, etc.), as well as the specific traits that your audience will be looking for, and try to demonstrate them subtly to the audience over the course of the talk via examples in your work. Consider ways to demonstrate your fundamental strengths as a scientist, such as the ability to question your methods and results to pursue deeper and more robust conclusions. If you have any particular successes on your record, such as big grants or markers of professional stature, don't be shy about mentioning them (but don't brag!). Having your publication citations and/or grants listed in smaller text at the bottom of corresponding slides is one way to show your accomplishments without explicitly mentioning them. Finally, you can casually highlight additional non-research skills (e.g., mentoring, outreach, collaborations) throughout your talk. For example, give credit to an excellent mentee who contributed to the data collection or to a gifted collaborator who added a component to your study. Your application materials likely included many of these things, but if you can find ways to incorporate them in your talk, a broader audience can see the full package of who you are.

Keep in mind Rule 1 (Know your audience) when deciding how best to showcase yourself, as different disciplines and subfields may vary in their perceptions of what makes a good scientist. For example, disciplines may vary in their appreciation for deep thought into specific mechanisms and experimental designs versus mathematical elegance and rigor. Others may prize applied over fundamental research or vice versa. This may be especially challenging if your research is interdisciplinary, so make sure to investigate what factors are valued most highly by the decision makers in the audience for your talk so you can design your talk to emphasize those aspects of your work.

## Rule 3: Impress the in-crowd...

Likely there will be people in the audience who work in the same field as you. Make sure to impress these experts with your knowledge and convince them you are worthy of being their colleague. You want to show them you have the sophistication and skills necessary to tackle advanced problems. Therefore, it's a good idea to do at least one "deep dive" during your talk in which you include one or two "muscle-flexing" slides. By this we mean slides with technical content that the general audience member may not be able to fully understand but for which you can flex your intellectual muscles and showcase your skills. Importantly, do not bluff or bluster in this section—a technical error in your deep dive would be fatal.

These deep dives shouldn't be long, or you risk losing most of your audience. However, a glimpse into the more advanced aspects of your work will convey that you're able to play in the big leagues in your field. Just make sure to reengage your audience after this show of prowess, ideally providing a big-picture summary of what you've just shown.

## Rule 4:. . . but also appeal to the out-crowd

In addition to impressing the specialists in the audience, you want to make sure the people who work outside your discipline are able to follow and enjoy your presentation. When preparing your talk, consider how you can present and frame the material so that even audience members from far-flung disciplines are engaged and can appreciate the broader relevance of your presentation. Be attuned to the breadth of the department you're visiting, as this can present various communication challenges. The diverse interests of faculty in a broad department (e.g., biology) can make it difficult to make your research program appealing to everyone. However, it can also be difficult communicating to a more focused department (e.g., molecular genetics) if your research is not exactly in line with what everyone else does. It helps to summarize the important findings of your research as you present them, in addition to their implications and why they are exciting, in case not everyone followed the technical aspects of your results. You can also make it easier for audience members from other fields to follow your talk by avoiding excess jargon and keeping your messages clear.

Emphasize the themes in your work that relate to the job and department you're interviewing for. If applicable and appropriate, it can help to subtly highlight connections between your research to research of other members of the department who have different specialties. But be careful not to overdo this, as it can become distracting.

## Rule 5: Play the hand you've got to optimal effect

Strategic choice of topics to include in your talk from among your entire research portfolio is critical for giving an effective and memorable job talk. Depending upon what career stage you are in (just finished PhD, postdoc, assistant professor, etc.), you may have a smaller or larger research portfolio. For an hour-long job talk, it is unlikely you will be able to effectively discuss everything you have ever done. And that's okay, because that is what a CV is for!

For your job talk, you need to assess your portfolio of published work, unpublished but completed work, and ongoing projects to determine which projects showcase your work most effectively and best match what the department is looking for in a future colleague. The most effective talk structures we observed were ones that focused on 2–3 research studies and that combined higher-level information with a few "deep dives" into the nitty gritty of a particular study (Fig 1). This talk structure will help you satisfy Rules 3 and 4 above, which discuss how you want your whole audience to understand and appreciate your talk, while also presenting the "meat" of your research and impressing those most familiar with your field. If you feel that this design doesn't convey the breadth or quantity of your productivity, consider adding a slide or two on the conceptual structure of your full research program in which you can show (with all your best citations) how all the pieces fit together.

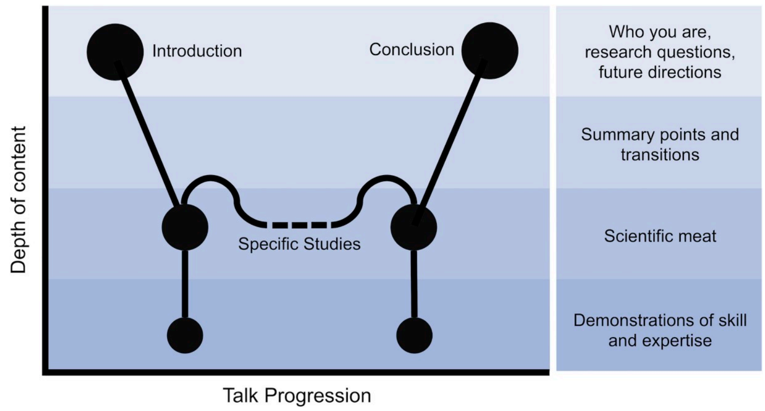



**Fig 1. Depth of content throughout the talk.** You want to start broad during the introduction to get everyone on board and then go into more depth on a few specific studies, including some "deep dives" to show off expert knowledge. Finally, you want to conclude your talk on a broad scale similar to your introduction. The dashed lines indicate flexibility in how many specific studies you incorporate into your talk, based upon your own research portfolio.

https://doi.org/10.1371/journal.pcbi.1007163.g001

In addition to presenting on your past and ongoing research, you need to clearly articulate your plan for your future research program. Tell the audience (and your potential future colleagues!) about your vision for your research lab both in the immediate future (next couple of years) and in the long term (5–10 years from now). This should also help differentiate you and your research from your previous mentors and their research programs. A critical part of establishing and maintaining a research program is your ability to generate funding. If you have already secured funding for your future research plans or you have a track record of successfully acquiring funding, then this is a great opportunity to bring this to your audience's attention. If you don't have independent funding yet, you can still demonstrate your awareness of the funding landscape and which funding opportunities are likely to support your research program. For example, in your future directions section, you might briefly touch on how one (or more) of your research questions aligns well with promising funding opportunities in your field, such as open research grants.

In organizing the structure of your talk and your transitions between topics, strive for a cohesive narrative that will make your talk more enjoyable to follow and easier to recall afterwards. What's the progression of your research? How did one study lead to the next, and what shaped your decisions about how to proceed? What ideas do you have for future research at this new job? Telling a story is always a great way to keep your audience engaged and makes your science more memorable.

## Rule 6: Give a good talk

A classic early paper in this series [5] provides ten useful rules for giving a good presentation. Read it! Showing you are a competent oral communicator is a vital component of giving an academic job talk. In addition to the universal suggestions from [5] (such as practicing for fluidity without over-rehearsing, making eye contact with the audience, and being enthusiastic and excited about your work), there are a few other pointers to bear in mind for a job talk. First, be aware that your job talk will be judged as an indicator of your ability to teach. Teaching is a crucial element of most academic jobs, but interview schedules often don't allow time to address it explicitly, so this doubles your incentive to give a clear and engaging presentation. Bonus points if you are able to expand people's understanding of technical aspects of your work—for instance, with a lucid explanation of your deep dive. Second, the job talk is a direct measure of your ability to sell your work and to act as an ambassador for the department in your future speaking engagements. Third, Rule 4 from [5] is "Make the take-home message persistent," and this is a particular priority in the swirl of an academic search in which four or five candidates may visit over the span of a few weeks. We found that a strong thematic structure, including outline and summary slides, was an effective way to emphasize and reiterate your key points and make them memorable for the audience.

Our next three pointers are more pragmatic, but they are still useful to consider. First, be sure to ask for guidance on talk length if you're unsure. For an hour-long seminar, the actual presentation length is typically 45–50 minutes, allowing for the fact that your host may burn precious minutes introducing you, and being certain to leave time for questions. Second, you should also make sure you understand the audiovisual equipment setup in the room where you are giving your presentation. If there isn't seminar preparation time on your schedule, ask for it! This way, you can ensure your presentation is loaded properly, your presentation slides appear how you expect, and you are able to navigate through them without glitches. It is a good idea to save your presentation in multiple formats in case you encounter compatibility issues with the primary format (e.g., if your presentation is in PowerPoint, also save a PDF backup version). Third, don't give your presentation while hungry. You want to exude energy and confidence, which may be difficult if you give a seminar later in the afternoon after many meetings and haven't eaten since lunch—so take note of your schedule and, if necessary, bring a snack to revive your energy levels before your talk.

The pragmatic pointers we mentioned are great for planning ahead, but overall, you should be adaptable. Problems can arise unexpectedly, and it's possible you'll be delayed by interruptions or a lengthy introduction. Do your best to not get flustered, to handle yourself with grace, and to end your talk on time. Make a note of places in your talk where you can go into greater depth if you're running ahead of schedule or places (particularly toward the end) where you can skim over the details more quickly if you're behind schedule.

## Rule 7: Be kind to your audience's eyes

Your slides should enhance your presentation, not distract from what you are saying. Make sure your slide aesthetics are appealing to the audience. Your slides should be clear and concise, without too much text. When you have text-heavy slides, you lose some proportion of your audience's attention while they read the text instead of listening to your words. So only display text that emphasizes the key points you will say out loud. Also, since the figures and images you present are especially important, you will want to construct figures specifically for your slides, keeping in mind that formatting for a presentation is typically different from formatting for a published paper. Refer to Box 1 for additional advice on qualities of good slides and common mistakes to avoid. You should also check out [5,6] for additional advice, noting that the rules in [6] are not specific to figures for presentations.

## Rule 8: Embody the future

Remember that you are the exciting next generation of scientists! Make sure to share your enthusiasm and your fresh ideas for research. Emphasize how your work is new and innovative, whether by showing new solutions to old problems or by describing ways to approach problems that have only recently been recognized. If appropriate, highlight how you will harness the latest technologies and methodological developments to advance your research. This will get the audience thinking about applications to their own research programs and how you'd be a valuable colleague to have around.

## Box 1. Qualities of good slides versus slide qualities to be avoided

Slide qualities to aim for:

- Good content:
    - Minimal text.
    - Figures that are readable and easily understood.
    - Figures created specifically for talks (rather than pulled directly from a paper). Talk figures are generally simpler than figure panels from a paper, with fewer items per plot, a focus on the key points, larger labels and axes, etc. Avoid having to tell your audience to ignore parts of the figure by remaking the figure without extraneous information.
    - If you have a complicated figure, you can animate your slides to build up the complexity as you explain it to the audience. For example, you can start by showing only a very simple plot and then layer on additional pieces of information as you explain them.
- Good design:
    - Clean background.
    - Consistent design throughout the talk.
    - Color-blind-friendly color palettes or alternative ways to distinguish differences on figures besides just color (e.g., using dotted versus solid lines to represent different measures in a plot).
    - Simple visual markers (silhouettes or clip art) that link ideas across slides and jog your audience's memory (e.g., a human silhouette next to parameters estimated from human data and a mouse silhouette next to data estimated from mice).

Slide qualities to avoid:

- Too much text.
- Text that's too small to read or overlaid on an image so that it's not legible.
- Busy background (e.g., photograph) that distracts from the text and/or figures you're showing on the slide.
- Figures with no or unreadable axis labels.
- Poor color combinations, including combinations that are difficult for color-blind viewers to make out (e.g., red/green, blue/green).
- Visual markers that don't convey any meaningful information, such as changing fonts and background colors. Even minor inconsistencies are distracting and convey a lack of attention to detail.

You can also emphasize other forward-looking traits you would bring to the job. Maybe you have developed a new online resource or are using a new mentoring or teaching style that helps make research more broadly accessible for students. Find ways to showcase how you are moving science forward and how you'll be a dynamic force for years to come.

## Rule 9: Don't blow it in the question-and-answer session

You're almost done with your job talk, so don't blow it during the question-and-answer (Q&A) session! You want to leave your audience with the best final impression and show that you can think and speak clearly in unscripted moments.

Here are some tips for a strong finish. When someone asks you a question, it can be helpful to paraphrase the question before beginning your answer. This gives you some extra time to compose your own thoughts and make sure you understood the question and ensures the rest of the audience hears the question. Regarding your actual responses, one cardinal rule is to never bluff. If you don't know the answer, you can say so, but then show how you would think through the question, or relate it to something you have done or know about. If somebody voices a fair criticism, then acknowledge it and discuss approaches to addressing it. If you can, convey enthusiasm in this situation—if it's truly an idea you've never considered, then treat this as an exciting and valuable scientific exchange, not an oral exam you are failing.

Remember that your audience likely includes people from outside your area of expertise, so it is possible you will get questions that seem to have missed key ideas from your talk. As with all questions, make sure you understand what the questioner is asking, and then take advantage of the opportunity to address any misunderstandings in a respectful, productive way. This is a great chance to demonstrate your ability to explain concepts clearly and concisely.

If there are predictable follow-up questions to your presentation, it can be helpful to have a few extra slides prepared. For example, if you presented a mathematical model using a schematic diagram, you may want to have a backup slide that shows the actual equations in case someone asks for more detail. If there is an extra data set or analysis that you'd love to include but just don't have the time, then a spare slide or two might enable you to deliver a home-run response if you get asked the right question.

Finally, remember to handle yourself with grace during the Q&A session. Be poised, calm, and respectful, and demonstrate your intellectual maturity—all of these are qualities people admire and are seeking in a future colleague. Another past article in this series gives rules for building your scientific reputation [7]; Rules 1, 2, and 3 are useful during both the Q&A session and the whole interview process! Which brings us to Rule 10.

## Rule 10: Be professional

Throughout this whole process, remember you are asking the host department to hire you as a (hopefully) long-term colleague in a small, tight-knit unit. Therefore, it is important to present a good image of yourself. You should dress appropriately for your job talk (i.e., not too casually). Even if you end up being a bit overdressed, it is better to leave that impression rather than showing up underdressed and being remembered as not having taken the job talk seriously. Be conscious of your body language and use of slang throughout your job talk and in any interactions you have during your visit. Humor can be a wonderful way to humanize and enliven your talk, but don't overdo it, and steer well clear of anything potentially offensive. While you are answering questions, or if you happen to be interrupted during your talk, remember to show yourself in the best light by being polite and calm, even if an audience member is being confrontational or rude.

You are an amazing and productive scientist (you wouldn't have been invited to give a job talk if you weren't!), but it's important to be clear about your specific contributions to the various research projects you present, particularly when the research is part of a big collaboration. It's essential to acknowledge your collaborators, especially junior mentees. This shows your audience that you are ready to mentor undergraduates, graduates, postdocs, etc., and most importantly, that you do not take collaborators' contributions for granted or claim them as your own. It's also good practice to acknowledge relevant previous work that your research and ideas are building upon, as you never know who is in your audience, and you don't want anyone to feel you are uninformed about or taking credit for this prior research. Again, you're asking to be hired into an academic family, and you want your new family members to be comfortable and excited about pursuing new research opportunities with you.

Finally, it is a nice touch to write thank-you notes after your visit (but see Rule 10 from [3] for an alternative opinion). These notes can be sent by email within a few days after the end of your job interview. How many you send is up to you, but we suggest sending follow-up notes to at least the search chair and the other key players in your interview visit. And don't forget about all the people who helped coordinate the logistical details for your visit!

In summary, the academic job talk is unlike most other seminars in its goals, context, and aspects of its execution. We have outlined some rules to help you put your best face forward in the job market (and to help all of us get the most out of the job search experience). There are additional resources online (e.g., [8] and [9] as two examples), and people should glean whatever insights they can from these sources. So do your preparation, nail the talk, and go get that job!

## Acknowledgments

This paper arose from discussions in a graduate seminar course jointly led by KEL and JOL-S. We thank other participants in the course including Katie Gostic, Natalie Lozano, and Bernard Kim for their thoughts on these topics.

## References

1. Tomaska L, Nosek J (2018) Ten simple rules for writing a cover letter to accompany a job application for an academic position. PLoS Comput Biol 14(5): e1006132. https://doi.org/10.1371/journal.pcbi.1006132 PMID: 29851981
2. Bourne PE (2011) Ten Simple Rules for Getting Ahead as a Computational Biologist in Academia. PLoS Comput Biol 7(1): e1002001. https://doi.org/10.1371/journal.pcbi.1002001 PMID: 21253560
3. Bourne PE (2014) Ten Simple Rules for Approaching a New Job. PLoS Comput Biol 10(6): e1003660. https://doi.org/10.1371/journal.pcbi.1003660 PMID: 24967974
4. Botham CM, Arribere JA, Brubaker SW, Beier KT (2017) Ten simple rules for writing a career development award proposal. PLoS Comput Biol 13(12): e1005863. https://doi.org/10.1371/journal.pcbi.1005863 PMID: 29240828
5. Bourne PE (2007) Ten simple rules for making good oral presentations. PLoS Comput Biol 3(4): e77. https://doi.org/10.1371/journal.pcbi.0030077 PMID: 17500596
6. Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLoS Comput Biol 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833 PMID: 25210732
7. Bourne PE, Barbour V (2011) Ten Simple Rules for Building and Maintaining a Scientific Reputation. PLoS Comput Biol 7(6): e1002108. https://doi.org/10.1371/journal.pcbi.1002108 PMID: 21738465
8. Reis RM. Giving a job talk in the sciences. 30 March 2011 [cited 2019 May 15]. In: The Chronicle of Higher Education [Internet]. Available from: https://www.chronicle.com/article/Giving-a-Job-Talk-in-the/45375.
9. Aguilar SJ. Tips for a successful job talk. 10 January 2018 [cited 2019 May 15]. In: Inside Higher Ed [Internet]. Available from https://www.insidehighered.com/advice/2018/01/10/advice-giving-effective-job-presentation-opinion.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules when considering retirement

**Philip E. Bourne** *

University of Virginia, Data Science Institute, Charlottesville, Virginia, United States of America

* peb6a@virginia.edu

OPEN ACCESS

**Citation:** Bourne PE (2018) Ten simple rules when considering retirement. PLoS Comput Biol 14(10): e1006411. https://doi.org/10.1371/journal.pcbi.1006411

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** October 4, 2018



**Funding:** The author received no specific funding for this work.

**Competing interests:** The author has declared that no competing interests exist.

You know a field is maturing when its early proponents think about retiring or, alas, pass away. So it is with computational biology. At 65, I think about retiring more. Not so much about retirement per se but in terms of what I want to accomplish before I retire and what retirement means to me in the first place. In other words, retirement is complex, and these ten rules probably (and hopefully) just start a discourse. For most scientists, including computational biologists, it's not a situation of now that you're 65, please accept your plaque, exit stage left, and goodbye. The questions I'm now considering are far more varied and nuanced. It's about where I want to focus my life and my energies. How much do I want to continue to mentor younger folk? How do I keep ties with colleagues I like and respect? How can I give back as much as possible to society and a profession that has treated me so well?

The focus here is on retirement as an option, not a requirement. After all, emeritus status in academia, government, and industry can typically go on in some form indefinitely.

It must also be said ahead of the rules themselves that in drafting the rules and having them reviewed by those acknowledged below, as well as discussing them with colleagues, friends, and family, that I was entering a very personal space. Efforts to coopt coauthors, which I like to do to provide a broader perspective, resulted in something to diffuse. How individuals think, or indeed choose not to think, about retirement is very personal. As such, the rules are more personal than I would normally write or have written on other subjects. Therefore, it may be that these rules do not resonate with you directly, but I hope they will at least make you think about what retirement means to you.

## Rule 1: Consider what you want to do when you get up in the morning

This would seem key and well covered in the endless retirement literature. This defines your new days. Currently, I rise at around 6 AM, make coffee, and then start on e-mail, preparing a talk or class or editing a paper or grant for a couple of hours before starting my official day. I have done that for many years. What would I fill my days with to replace that? Would I keep doing some of it? Would I sleep in, which I have not done for years? Would I read, go out walking, go motorcycling, activities I do now but want to do more of? Would I start new activities?

My current thought is to start new activities before any major life change with enough vigor to determine whether I like them or not. It seems to me that starting a new lifestyle overnight could be a mistake; such drastic change would be jarring. Instead, it seems sensible to experiment with different options, figure out what you like and spend more time on them leading up to retirement. For me, that means getting a fix-me-up vintage motorcycle, learning at least one new programming language to be as proficient as my graduate students (unlikely), reducing the backlog on my Kindle, and spending more time with family. Inherent in all these desires is the notion of not biting off more than one can chew. For if that is the case, it is business as usual and hence no sense of retirement at all.

## Rule 2: Consider the financial and scholarly implications

If I have to generalize, when looking at my colleagues in industry, government, and academia in many parts of the world, money is, thankfully, often not the major factor in considering if and when to retire. They have frequently been working for a considerable period of time at institutions that have good retirement plans. Moreover, in keeping with the flexible notion of what "retirement" means, there may be opportunities for paid part-time work not considered previously, such as being a consultant.

If you are one of the few people who are young and reading this, let me state what you have heard many times I am sure. Plan your financial future now. The wonders of compound interest should not be ignored. Financial security will definitely be part of your retirement calculus and that day comes sooner than you think.

Without money as a major concern, pro bono activities are more likely to arise as an unpaid consultant, scientific advisory board member, and so on. These activities also allow one to keep up to date with the latest science if one chooses. Although that can be harder if one does not have access to the closed-access literature offered by universities, etc.—a sad statement. Maintaining an adjunct affiliation with an institution that has a closed access journal subscription can address this, but there are often strings attached like requirements to teach.

## Rule 3: Consider realistically what can yet be accomplished professionally

This could apply up until you retire but, based on Rule 1, could continue well into the "retirement" years. If you are obsessed with accomplishing more, then probably best not to retire at all. There is also the question of accomplish for whom? If it is for yourself, that might lead to a different path than if it is in service to others.

For me, I think about it in terms of whether I have achieved my career goals. In thinking deeply, I realize I never really had specific goals at the outset. Instead, they appeared as I lurched forward from researcher and added mentor, administrator, and institution builder. For me, it's about contentment—I feel content and that is what matters to me. What remains is to see things through as covered by other rules and to make my list of things yet to do. See Rule 8.

## Rule 4: Consider your loved ones

I have told this story to students many times as part of a lesson in life–work balance. When my son was little, he would sit on my lap and hold my head in his little hands and make eye contact. Some 12 years later, my daughter did exactly the same thing. My immediate reaction was to put it down to their shared gene pool. Then one day I had a horrible realization—it was nurture, or lack thereof. Both my children, at many different times, had come up to me and said, "Dad, help me with. . ." and I, without my eyes leaving the computer screen, would say, "ah ha, ah ha."" For years, after that realization and now that they are adults, I have said to them, "Did you feel neglected?" They always respond, "Nah, of course not, Dad." I am not so sure and that is something I want to address in the coming years. That I did not practice what I preach to my students is yet another matter.

Beyond that and in deference to my wife of 35 years is to redress the sacrifices she has made. The partner of a computational biologist, or any scientist for that matter, is one of sacrifice, both professionally and emotionally. One's partner has not only to love you but to love your work as you do. Significant redress is long overdue.

## Rule 5: Consider your colleagues and take steps before you retire

You are, of course, not obligated to continue working. However, there are many professional considerations that may affect the calculus: for example, if you do not retire, you may be preventing someone else from getting your tenured appointment; if you do retire, you may feel as if you are leaving students in the lurch, while leaving the courses that you teach and various administrative duties to fall to some poor soul and so on. Some of what you impart to colleagues is unique and will be lost but that must be balanced against your own wishes and needs.

Try and mitigate what will be lost by nurturing others to take over those responsibilities ahead of time. Of course, this does not just apply at retirement but also when you change jobs or responsibilities. The most important management and leadership skill I have learned over the years is to plan for your successor the day you start a new position. This makes separation smoother.

## Rule 6: Consider what your field will gain or lose and then ignore it

I am not attempting to speak for anyone but myself, but I think that it is far too presumptuous to think I am contributing anything unique. Over my 45-year career, science in general, and computational biology in particular has become a team sport. The complexity of the work demands no less. We are each part of the solution but not uniquely so. From the beginning, Isaac Newton's notion about "standing on the shoulders of giants" rings true. As time passes, individual contributions fade into the fabric that is science past, and only in rare cases will individuals stand out in the eyes of future generations. I imagine that there are not a great many papers from the seventies or scientists in your field that you can cite or name from that era. Be satisfied to be part of those broad shoulders upon which future scientists will stand.

## Rule 7: Consider your health and wellbeing

This is a no-brainer, ill-health can strike at any time. Less likely if you have maintained a regime of exercise, good diet, enough sleep, etc. over your lifetime. One question to ask is, if something happens health-wise and I am still working, will I have lost the opportunity to do other things that I will now not be able to do. Stated another way, will my final words be, "I wish I had written just one more paper" or something else. Planning around that scenario would seem wise.

## Rule 8: Consider opportunities not available otherwise and be prepared for the unexpected

As scientists, we are inquisitive and life-long learners, yet we never have enough time for all the avenues, professional or otherwise, we wish to explore. For me, those things include one or two reviews that I think should be written, an autobiographical perspective on biomedical data (seriously), and a small number of research threads still to be pursued. All can be done at a different pace and likely with a different perspective than when heavily involved in running a research laboratory and institute. What's more, they can be done easily. What may plague one during a career as a computational biologist—the ability to work at home as easily as working in the lab—now can become an asset. You can easily keep your hand in but at your own pace and by doing those tasks that one enjoys without the need to write grants or undertake a myriad of administrative tasks.

Beyond that, those retired or at least slowed down have told me that completely unexpected pursuits arise, either professionally or otherwise. I am excited to see what interesting opportunities arise and am ready to embrace the unexpected.

## Rule 9: Consider the possibilities of giving back in new ways

This could take many forms and be rewarding. Examples that come to mind are various types of voluntary work, either working for or starting a foundation and mentoring at various levels. There is always a need for free expertise in science, technology, engineering, and medicine (STEM). In fact, refer to Rules 1 and 2: consider how giving back could factor into your day-to-day retirement lifestyle.

## Rule 10: Put your scholarly legacy in order

Part of your legacy is what you leave in the scholarly literature (noting Rule 6). Much of that is already in order thanks to our journals and practice of publishing. Yet there is more to one's scholarship, some of which might prove to be more useful than the papers themselves. There are data, software, laboratory protocols, course materials, and so on. Are they available and in the best shape for reuse or for passing knowledge forward? At the point of retirement, presumably, one is less concerned about getting credit for hot new science and more about ensuring that knowledge is not lost. It is the liberating time in one's career to do what is right, not do what it takes to get promoted.

There you have it. Most of you probably did not get this far as you are not even close to reaching retirement, and it seems ridiculous to contemplate. Trust me, when you do get to read to the end, you too will be well aware that retirement comes quick enough. Whatever your later years bring, I hope they are as rewarding as I plan to make mine—someday.

If you did get to the end, perhaps you are retired and had the time and inclination to read these rules. Hopefully, you also have the time to comment on them from your own perspective, for as I said at the outset, first, these rules derive from someone who thinks about retirement but who has not yet retired, and second, everyone will have a unique perspective on the subject.

## Acknowledgments

Thanks to Fran Lewitter and David Searls for their perspectives on being "retired" and to Ira Klein and Teri Klein, who were kind enough to weigh in with corrections and their own unique perspectives.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for writing a cover letter to accompany a job application for an academic position

**Lubomir Tomaska[1]*, Jozef Nosek[2]**

**1** Department of Genetics, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovakia, **2** Department of Biochemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovakia

* lubomir.tomaska@uniba.sk



OPEN ACCESS

**Citation:** Tomaska L, Nosek J (2018) Ten simple rules for writing a cover letter to accompany a job application for an academic position. PLoS Comput Biol 14(5): e1006132. https://doi.org/10.1371/journal.pcbi.1006132

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** May 31, 2018



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

It is becoming a difficult task to find an academic position that is best suited for one's capabilities and preferences. In an extremely competitive environment [1], there are tens of applicants (and often more) per a single position. As a result, the hiring committees judge the quality of the candidates based on numerous criteria, including previous achievements listed in their CVs; recommendation letters from their instructors, supervisors, or peers; technical and presentation skills; and research plans. In many cases, the first encounter with the applicants is mediated by their cover letter. If well crafted, the letter can simultaneously act as an introduction, a first-stage filter, and a cogent, compelling argument for one's candidacy (e.g., [2–6]). On the other hand, a generic, boring, uninspiring cover letter full of typos will increase the probability of dismissal of the application, oftentimes irreversibly blocking the applicant's entry to a potential dream position. The list of rules below should be helpful in composing a cover letter that will serve as a catalyst for pushing the application into the next stages of evaluation. The provided rules are specifically designed for job applications for an academic position (e.g., PhD student, postdoc, lecturer, faculty member). Although most also apply for other types of jobs (e.g., in industry), these may have specific requirements that need to be taken into account.

## Rule 1: Before starting the application, make a list of pros and cons for the position

On 11 November 1838, Charles Darwin proposed to his cousin, Emma Wedgwood, and wrote in his diary, "The day of days!" What followed (beginning 29 January 1839) were 43 years of happy marriage, underlining the fact that Darwin's proposal and Ms. Wedgwood's acceptance were both correct decisions. Perhaps what is not so widely known is that Darwin, in the months immediately preceding his engagement, had written two notes weighing up the pros and cons of marriage [7]. Following Darwin's example, before starting to compose an application letter, it should prove helpful to summarize the reasons in favor of (and also against) applying for the position. Perform detailed research of the available positions that may suit you. Go beyond the information provided by job ads and institution websites; consult mentors, advisers, and colleagues. Start writing the cover letter only if the pros of application outweigh the cons. This will help you to compose a text in an enthusiastic style.

## Rule 2: Remember KISS—Keep It Short and Simple

This is one of the key concepts in effective speaking and writing. For a cover letter, if not required otherwise by specific formal requirements, two pages are the maximum. Address the letter to a named person. A short introduction serves as a "handle," i.e., it should get attention from the committee members [2]. Then you can proceed to a brief and clear summary of your most important—and relevant—qualifications. The balance between your research, teaching, and administration (or other) skills depends on the nature of the position. Next, explain what attracts you to the position and how it fits into your career plan. To be concise, stay focused. Anything less than a sharp focus and your readers will quickly lose interest and move on to the next application. Do not duplicate your CV. Rather, emphasize what does not get covered or rise to the surface in your CV or résumé. Make sure your cover letter is consistent with your CV. In the case of a digital cover letter, you can provide active links to information that may be relevant such as your website or list of publications (e.g., as a link to ResearcherID, ORCID, or GoogleScholar). The closing of the letter is as important as its opening. Do not let it meander to an indefinite or weak last paragraph. End your letter decisively by including a statement expressing interest in an interview.

## Rule 3: Be original, nonconformist, and personal

Although it is useful to read cover letters of successful candidates, do not get too influenced by their style and content. Be yourself. Think of your cover letter as the opener to your application, similar to a cork that represents an entry to the contents of a wine bottle. Just as a cracked or rotten cork will discourage a user from pouring the contents of the bottle into a glass and evaluating all of its attributes, an uninspiring cover letter might prevent the recruiter from reading the full application and assessing your suitability for the position.

## Rule 4: Show motivation and sincere interest

If you are applying for a research position, try to address the following questions: why you are choosing this position; what is exciting about the projects performed in the laboratory; and what part of the project you would like to pursue. You need to show that you did not just read the titles of recent publications but that you are familiar with the methodology, experimental design, and analysis as they are performed by your prospective employer. This will help you suggest future experiments or research directions aimed at better understanding of corresponding phenomena. In case of a faculty position, explain why you plan to pursue your career at the given institution; indicate how you will benefit from the collaboration with the groups at the department and vice versa, and how you plan to obtain support for your research; describe your experience with supervising students and postdocs; and show motivation for teaching, if the position requires teaching-related activities [8].

## Rule 5: Provide an honest description of yourself

Provide a glimpse of your personality, possibly in the form of a story that highlights your specific characteristics. Pick a few adjectives that describe you most of the time, regardless of the situation. You may provide information about the path that led to your interest in a particular field. Be positive.

## Rule 6: Highlight your strengths

How will the lab or institution benefit from having you onboard? List your major achievements and technical skills. State explicitly how your abilities and interests align with the position. Expand on what makes you especially suitable or appealing for the specific position you are applying for. Explain which of your strengths may set you apart from other candidates. If your background does not exactly match all of the criteria that the employer seeks, spell out what you are willing to do to learn the specific skills that the hiring organization needs (e.g., taking a special course or training). Do not forget about your behavioral strengths, such as your ability to lead a project, to work as a team member, or to be an effective communicator [9].

## Rule 7: Do not recycle

If you are sending applications for several positions, do not use the same text and just change the name of the institution. It is important to tailor your letter to the position you are applying for. Every research group and every institution prides itself on certain characteristics that make it unique. Therefore, generic, template-like letters are prone to be identified. They can make you seem rushed, noncommittal and not particularly interested in the position advertised and are most likely to end up in the recycle bin.

## Rule 8: Avoid overstatements and false claims

Overstatements tend to be annoying, and false claims, when uncovered, result in immediate rejection. Do not make yourself look better or more qualified. Avoid pompous metaphors and clichés. Be honest and truthful. Do not exaggerate. Also, do not be unrealistic in what you could achieve should you be offered the position, i.e., do not list promises that you cannot keep. Remember that it is always better not to be accepted for a position than to run into troubles or conflicts while holding it.

## Rule 9: Do not underestimate the formal quality of the letter

If a letter reads well, looks good, and is devoid of spelling and grammatical errors, then the reader will have a tendency to associate those qualities with the candidate. If someone is unable to express him/herself without errors, that triggers an immediate rejection [5]. Proofreading for content, accuracy, and style is crucial. Here are some suggestions: Spell check and get a colleague or trusted personal contact to check spelling and readability, too. Automated spell checkers may not catch inappropriate usage of similarly spelled words or homonyms; be particularly careful about writing the recipient's and institution's names correctly, check any dates and addresses you are referencing, avoid lists or bullet points, and avoid unusual or unreadable fonts.

## Rule 10: Plan ahead and do not rush

When you think the letter is in the best shape possible, try to put yourself in the position of the recruiter. To do so, set it aside for a couple of days. Such detachment will help you to see flaws that were not obvious at the time of writing. Format the letter even if it is being sent by email. Ask for references or recommendations in advance. Allocate sufficient time to finalize the letter. Note that different countries may have different application processes and that not everything indicated above will apply to every country. Therefore, have someone with knowledge of local customs review the letter as well. Finally, do not miss the deadlines.

## Conclusion

Although a well-written cover letter is only a small step towards a successful job application, it may provide an important advantage by sparking the interest of the hiring committee for that particular applicant. Even if it is not formally required, it should still be included with the application, as it can help to underline the applicant's qualities related to the position description and it may point to a particular section of a lengthy résumé where they can find more detailed information [5]. To put it metaphorically, a cover letter is a handle on the door to the application, and depending on its quality, the door can remain closed, or it can be opened.

## Acknowledgments

We thank Smaranda Willcox (University of North Carolina at Chapel Hill) for the English language editing. We also thank Annamária Hlavatá, Paulína Káňovičová, and Jakub Králik (Department of Genetics, Comenius University in Bratislava) for sparking the idea about this Ten Simple Rules article.

## References

1. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US Biomedical Research from its Systemic Flaws. Proc Natl Acad Sci USA. 2014; 111:5773–5777 https://doi.org/10.1073/pnas.1404402111 PMID: 24733905
2. Borchardt JK. Writing a Winning Cover Letter. Science Careers 2014. Available from: http://www.sciencemag.org/careers/2014/08/writing-winning-cover-letter. Accessed on 8 April 2018.
3. Gould J. Make your Cover Letter and CV Stand Out. Nature Jobs blog 2014. Available from: http://blogs.nature.com/naturejobs/2014/02/28/make-your-cover-letter-and-cv-stand-out/. Accessed on 8 April 2018.
4. Miller Vick J, Furlong JS, Lurie R. Academic Job Search Handbook. 5th Edition. University of Pennsylvania Press 2016
5. Ogden LE. Under the Covers. Nature. 2016; 538:129–130. https://doi.org/10.1038/nj7623-129a
6. Carr L. How to Write a Cover Letter for Academic Position. Available from: http://www.jobs.ac.uk/media/pdf/careers/resources/how-to-write-a-cover-letter-for-academic-jobs.pdf. Accessed on 8 April 2018.
7. Darwin Correspondence Project. Available from: https://www.darwinproject.ac.uk/tags/about-darwin/family-life/darwin-marriage. Accessed on 8 April 2018.
8. Howard Hughes Medical Institute and Burroughs Wellcome Fund. Making the Right Moves: A Practical Guide to Scientific Management for Postdocs and New Faculty. Available from: http://www.hhmi.org/sites/default/files/Educational%20Materials/Lab%20Management/Making%20the%20Right%20Moves/moves2.pdf. Accessed on 8 April 2018.
9. Dolan B. How to Write an Academic CV. Genes to Genomes–a blog from the Genetics Society of America 2018. Available from: http://genestogenomes.org/how-to-write-an-academic-cv/. Accessed on 8 April 2018.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for writing a career development award proposal

**Crystal M. Botham[1]*, Joshua A. Arribere[2], Sky W. Brubaker[3], Kevin T. Beier[4,5]***

**1** Stanford Biosciences Grant Writing Academy, Stanford University, Stanford, California, United States of America, **2** Department of Molecular, Cell, and Developmental Biology, University of California, Santa Cruz, Santa Cruz, California, United States of America, **3** Department of Microbiology and Immunology, Stanford University, Stanford, California, United States of America, **4** Department of Biology, Stanford University, Stanford, California, United States of America, **5** Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, United States of America

* cbotham@stanford.edu (CMB); beier@stanford.edu (KTB)

Check for updates

OPEN ACCESS

**Citation:** Botham CM, Arribere JA, Brubaker SW, Beier KT (2017) Ten simple rules for writing a career development award proposal. PLoS Comput Biol 13(12): e1005863. https://doi.org/10.1371/journal.pcbi.1005863

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** December 14, 2017



**Funding:** JAA was supported by an NIH F32 from NIGMS (5F32GM112474-02), SWB was supported by an NIH T32 from NIAID (5T32AI07290), and KTB was supported by an NIH K99 (K99DA041445) from NIDA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Career development awards support intensive training and provide valuable funding for promising postdoctoral and/or junior faculty–level scientists during the transition to research independence. Therefore, career development awards are typically competitive and are a significant undertaking. Many countries have such awards, with the National Institutes of Health (NIH) K series being the most well known in the United States. If funded, the career development award demonstrates success in obtaining research funding, a key factor for future success. In addition, regardless of whether or not the application is funded, a well-designed proposal will help launch the transition to independence by providing a detailed plan for career development and future research. This guide provides an overview for formulating a strong career development award, with principles generalizable to other grant proposals.

## Rule 1: Give yourself time

As a guideline, we recommend that 3 months of full-time effort be dedicated to completing the application. During this writing time, you may have limited bandwidth for your research and other activities. This may seem excessive, but career development award applications often require more documents than senior-level grants. Even for experienced writers, soliciting and incorporating feedback on the various documents can be a time-consuming process. We recommend that you start on your application as early as possible and work steadily.

First, read all the instructions and create a list of all the required documents. Create a realistic timeline that outlines when you will draft, elicit feedback on, edit, and finalize the various documents. Prior applicants can be a particularly valuable resource—they can provide a sense of the required time commitment and may be willing to share their own proposal and reviewers' comments to show you how the proposal was evaluated. Ideally, try to obtain copies of previous successful applications. For NIH K series awards, one particularly valuable resource can be the NIH Research Portfolio Online Reporting Tools, Expenditures and Results (RePORTER) tool [1]. On this site, you can read abstracts of successful grant applications. In addition, if you don't personally have access to copies of funded grants, you can consider reaching out to postdocs at your institution that you identify through the NIH RePORTER website.

## Rule 2: Write using the review criteria as your guide

The reviewers of your application are busy. They are running labs, teaching, mentoring, writing papers and grants, working in the clinic, serving on committees, and/or balancing other personal obligations. Reviewing grants doesn't replace these responsibilities but rather gets added on top of them. As a writer, strive to make reviewing your proposal as quick as possible by making it easy to read (see Rule 6) and easy to score.

To evaluate your application, reviewers are provided a set of criteria, which is often available to applicants. For example, the NIH K99/R00 lists the scored review criteria, including specific questions for reviewers, in the Program Announcement, Section V (Application Review Information). This section articulates the funding agency's expectations. It is therefore essential that you provide direct, clear answers to these questions in your proposal. You can address these questions at multiple points in your proposal and emphasize your answers with unique fonts (*italics*, **bold**, or underline). Furthermore, you can use the review criteria as a source of content, structure, and language to write your proposal.

As an example, consider that the reviewer is asked the following question (emphases added): "has the candidate presented strategies to ensure a robust and unbiased approach, as appropriate for the work proposed?" In this case, you can consider including a sentence in your Research Plan that uses the same content, structure, and language; e.g., "*these multiple strategies provide a robust and unbiased approach to answer my research questions*." Italicizing the sentence will make it stand out on the page. This parallel structure ("strategies," "robust," "unbiased") makes it easy for your reviewer to locate and answer that review question. If the reviewer needs to search for or infer the information, it may negatively impact your score.

## Rule 3: Present a plan to differentiate yourself from your mentor

Reviewers for career development awards will consider whether the proposal offers a clear plan for separation from your mentor's research. First, initiate a conversation with your mentor about which aspects of your future work you can take to your own lab. Make sure you and your advisor agree on this plan. Then, make sure your documents, as well as your mentor's support letter, articulate a consistent plan for the mentored and independent phases of your career.

Aim to demonstrate your plan for differentiation from your mentor at multiple points throughout your proposal. For example, in your NIH "biosketch," you can describe how the combination of your past and proposed research gives you a perspective that is unique from your mentor's. If you learn new techniques from collaborators outside of your lab, state how this will give you a niche distinct from that of your mentor.

## Rule 4: Elicit feedback on your Specific Aims early and often

The Specific Aims document is the executive summary of your research proposal and the single most important document of your application. In fact, it is likely the first document your reviewers will read. It is therefore crucial that your Specific Aims document is clear, engaging, and exciting.

Often limited to 1 page, the Specific Aims document must succinctly convey (1) your proposal's importance, (2) what you will do in the appropriate amount of detail, and (3) how the proposal will contribute to your training. Most of all, it is imperative that your Specific Aims document ignites the reviewers' desire to read more.

We recommend reading Yuan et al., 2016 [2] and the Specific Aims chapters in *The Grant Application Writer's Workbook* [3] for tips on how to construct your Aims page. Then, construct a preliminary outline that includes the key information you want to convey. Ask your mentors, colleagues, and friends for feedback on the overall framework of your ideas. Seek feedback from an audience most similar to your reviewers because this will help tailor your proposal for your target audience. Next, write an early draft of your Specific Aims document and again seek feedback by requesting 3 to 4 prioritized comments from each reviewer. This will help you address the most important concerns. An iterative process of eliciting feedback and revising will greatly strengthen your writing. Getting feedback from both experts from your field (e.g., within your lab) and nonexperts (outside of lab) is important.

Once your Specific Aims document begins to take shape, engage your funding agency. Present it to the program officer (read "What to Say—and Not Say—to Program Officers" [4]) or other relevant official to check whether your proposal is suited for the agency's funding goals.

## Rule 5: Create a Research Plan that bridges the gap between a scientific unknown and an expected payoff

Don't fall into the common trap of simply listing the Aims and experimental details in your Research Plan. This document should describe how your project will bridge the gap between something needed and an ultimate payoff. Therefore, the need and the payoff should be clearly stated at the onset and reiterated (sparingly) throughout the document. It's your job to convince your reviewer that your plans (Aims) are bound to successfully bridge that gap you wish to fill. You can use the following questions to help guide your Research Plan and convince a reviewer that your research is important and feasible:

1. Why is the project needed?
2. What is innovative about the project?
3. How will the project be completed?
4. How long will the project take?
5. What are the expected payoffs from the project?

Answers to the above questions correspond to the general subsections of the Research Plan, as follows: (1) Background & Significance, (2) Innovation, (3) Research Approach, (4) Timeline, and (5) Conclusions & Future Directions.

## Rule 6: Use clear writing

Clear, concise writing creates stronger and more memorable arguments. While there is a temptation to write up to the page limit of a document, a structured argument supported by only the necessary information will be more compelling.

Keep it simple. First, break down each required document into its most essential components (see Rule 2 for how to use the review criteria as your guide). Consider making bullet sentences for each point or using just a few words to state your argument. If this is not sufficiently simple, it needs to be refined. This is a great opportunity to discuss your ideas in their simplest form with mentors and colleagues.

Next, slowly build it back up. Connect the bullets with appropriate transitions so that each point flows logically to the next. Think of it as holding your reviewer's hand as he/she reviews your documents. You must lead him/her to what needs to be done next.

Lastly, polish your writing by using sentence structures that are clear and concise. Some samples are shown in Table 1.

**Table 1. Suggestions for how to implement clear writing.**

| Suggestions for clear writing | Instead of: | Write this: |
|---|---|---|
| **1. Limit use of the verb "to be."** | X is an indication that Y | X indicates Y |
| **2. Limit prepositional phrases.** | The instrument in the lab is necessary | The lab instrument is necessary |
| **3. Use direct, active-voice sentences.** | Provides justification for | Justifies |
| **4. Avoid noun forms of verbs (nominalizations).** | The application of these techniques can | Applying these techniques can |

For more information about clear writing, see Dr. Kristin Sainani's *Writing in the Sciences* course [5].

https://doi.org/10.1371/journal.pcbi.1005863.t001

## Rule 7: Construct a strong mentoring team

Career development awards are mentored awards. Therefore, they frequently require a dedicated team of mentors that will guide and evaluate your progress during the proposed funding period.

An appropriate mentoring team should have the necessary research qualifications, experience, scientific stature, and mentoring track record to help you enact your research and training plans. An advisory team typically consists of your primary mentor, an optional co-mentor, and 2 to 4 additional mentors. Select a team that complements your strengths. For instance, if you are proposing to use a technique new to you, find an advisor who has published using that technique. Clearly describe in your proposal how this relationship will enable you to learn the technique. Obtain letters from your mentors that affirm their commitment to mentoring you and describe, specifically, how they plan to foster your development and transition to independence. Be sure that letters from your mentors are very detailed. Include information such as how often you will meet with the mentor, how long these meetings will be, whether you will attend lab meetings, and how progress will be assessed.

As with other documents, these letters need to consider the review criteria and the funding opportunity. It is useful to meet with your mentors and advisors to ensure a mutual understanding of your research and training plan and to guarantee that they clearly articulate their roles in any provided letters. You may consider providing a template that includes the information that you wish them to convey.

## Rule 8: Design a career development plan to equip you for independence

Unlike most lab grants, career development awards include a training component. Use your career development documents to chart a path to your future success by addressing the following questions, sequentially narrowing down from big picture to specifics:

1. What are your career goals?
2. What skills do you need to achieve your career goals?
3. What activities will you engage in to attain those skills (include course numbers, meeting titles, and the names of any principal investigators [PIs] involved)?

Be thorough and rigorous: include a timeline to achieve your goals, establish a plan to monitor your progress, and obtain letters of support from additional PIs as necessary (see Rule 7).

Critically, you must articulate why additional training is necessary. If you struggle to articulate the need for additional training, you may already be equipped for your future career and may not need additional training.

## Rule 9: Use prior training and experiences to highlight your potential

Many career development awards require that you describe your background and career goals. Furthermore, you may be required to submit a document like the NIH biosketch that summarizes your prior training and experiences. Your reviewers will use these documents to determine whether you have the potential to become a successful, independent investigator based on what you have done to date and how that relates to what you propose to do. Therefore, you must demonstrate your research productivity and the quality of past training.

In these documents, instead of providing a laundry list of publications or research expertise, use elements of the classic story arc if possible. First, outline the historical background that inspired the work, including the challenge or knowledge gap addressed. Next, detail your contributions as well as the central findings and any expertise you gained. Lastly, describe the impact of your research and why it was significant.

You may also need to write a personal statement. Use this opportunity to tell a story that describes why you made key career choices and provide evidence for your long-standing and continued commitment to research. It is helpful to conclude this subsection by describing how your prior activities in conjunction with your proposed training will enable you to achieve your career goals.

## Rule 10: Weave a consistent story throughout all documents

Maintaining continuity is a very important but often underappreciated aspect of a successful grant application. All aspects of the proposal, from supplemental documents to recommendations, should mutually reinforce one another. Do not assume that saying something once is sufficient—if something is important, state it many times, in many documents. For example, if you plan to learn a technique from a member of your advisory committee, highlight this in your research strategy, sections concerning your training and career goals, as well as in letters from your mentor and the specific advisory committee member. If you state that you will visit a collaborator's lab, you should budget related travel costs into your proposal. The more you link the different components of your research and career training plan together, the more your reviewer will feel that the proposal was thoroughly planned and well developed. Also, your reviewer will have an easier time finding relevant information (see Rule 2). Allow for time within your writing schedule to synchronize components of your proposal to tell a single, coherent story.

## Conclusion

These 10 simple rules provide a framework to construct your career development award application. The process may take a significant amount of time and effort, but it's worth the potential payoff. Regardless of the award outcome, it will be time well spent planning your research and career. Lastly, here is the most important thing to remember: apply! You can't win if you don't play.

## References


1. NIH RePORTER. https://projectreporter.nih.gov/reporter.cfm. Accessed on 20 August 2017.
2. Yuan K, Cai L, Ngok SP, Ma L, Botham CM. Ten Simple Rules for Writing a Postdoctoral Fellowship. PLoS Comput Biol. 2016; 12: e1004934. https://doi.org/10.1371/journal.pcbi.1004934 PMID: 27415752
3. Russell SW and Morrison DC. The Grant Application Writer's Workbook–National Institutes of Health Version. www.grantcentral.com. Accessed on 20 August 2017.

4. Spires MJ. What to Say—and Not Say—to Program Officers. The Chronicles of Higher Education. 2012. http://chronicle.com/article/What-to-Say-and-Not-Say-to/131282. Accessed on 20 August 2017.
5. Sainani K. Writing in the Sciences. https://lagunita.stanford.edu/courses/Medicine/SciWrite-SP/SelfPaced/about. Accessed on 20 August 2017.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules in considering a career in academia versus government

**Philip E. Bourne***

Data Science Institute and Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America

* peb6a@virginia.edu



OPEN ACCESS

**Citation:** Bourne PE (2017) Ten simple rules in considering a career in academia versus government. PLoS Comput Biol 13(10): e1005729. https://doi.org/10.1371/journal.pcbi.1005729

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** October 5, 2017



**Funding:** The author received no specific funding for this work.

**Competing interests:** The author has declared that no competing interests exist.

The author is the co-founder and founding Editor in Chief of PLOS Computational Biology. He spent 20 years at the University of California San Diego as a professor of pharmacology and for the last two years there as Associate Vice Chancellor for Innovation and Industrial Alliances. From 2014-2017, he was the Associate Director for Data Science at the National Institutes of Health and a Senior Investigator in The National Center for Biotechnology Information. As of May 2017, he is the Stephenson Professor of Data Science, Director of the Data Science Institute and a professor in the

This article is focused on a career point at which a higher degree is in hand—perhaps along with some practical experience—and it is time to make a career decision. One such decision might be between an academic scientific research career versus a non-research career in government service. There are many other opportunities, of course, and industry versus academia has been well covered previously in this series [1]. With federal research funding as limited as it is, early-career scientific researchers are increasingly looking at nonacademic pathways; government service is one option. An example choice might be between accepting a postdoctoral fellowship or tenure track assistant professorship versus becoming a program officer for a funding agency, working in government relations, or working in government policy development. Obviously, these are only a couple of the many career choices available in academia and government. These rules are meant to be as generic as possible by recognizing the broad similarities and differences that exist in the 2 work environments. The rules do not cover the obvious differences, such as the ability to teach in academia but likely not in government.

As indicated, academic research and government service both cover large amounts of career territory. While trying to be as evenhanded as possible between these 2 career paths, undoubtedly, bias stemming from my own experience creeps in, and it is important to understand from where my perspective derives. I have spent most of my career in academia as both a professor and a university administrator. More recently, I spent 3 years in the United States federal government, where I had both an administrative and research role, both in biomedicine. My experience is far from that needed to provide a complete picture of career options. For example, it does not address government service, federal or state, outside of the US. Nor does it truly address the myriad of options outside of working for a government funding agency focused on biomedical research. More problematic is having worked 3 years in government versus over 40 years in academia. Undoubtedly, it is a different article than if I had spent 40 years in government service and 3 recent years in academia. Keeping in mind these limitations and the fact that I have been strongly influenced by the excellent reviews of the first version of this article, what follow are the rules I have to offer, rules which are made as generic as I know how.

Remember also that career options are not for life, and experience in government can be very useful to furthering a career in academia and vice versa. This is something that I can attest to, and which I try to capture.

## Rule 1: Public good means different things

As an academic, I rarely thought about public good, defined as a commodity or service provided without profit to all members of society. Yes, I did my research with the idea of

Department of Biomedical Engineering at the University of Virginia.

improving the human condition, but that was about it. I gave little thought as to how efficiently and productively I was using public money and the impact that was having on the public. I considered myself central to the enterprise. In government service, the enterprise is central. In government, you are part of a much bigger collective enterprise than the individual research laboratory and its associated discovery. Much of what follows flows from this notion. If you read no further, this defines what is fundamentally different, and it really does take 2 very different personalities. Personally, I would not have been happy in government service if I had not first had an academic career with which I was satisfied. I needed to satisfy the individual before the collective. This sounds selfish, and in some ways it is. This is a good reason, if you are in academia, to respect those in government service around you for their unselfish work! Thinking about the individual versus the collective is also a good basis for really assessing your motives in considering one career path over the other—be honest with yourself.

## Rule 2: Visible rewards and recognition are different and likely on different timescales

In academia, highly cited papers, grants awarded, teaching awards, etc., define success. Again, it is very much centered on the individual or small group. In government, relatively speaking, a new policy, program, etc., likely represents the work of many people. Of course, as humans, we want to be recognized for our efforts. In government, such recognition is not citation and tenure but likely accolades from colleagues, service awards, or promotions. Academia is about individual reward; government service is more about collective reward. Academia is more about broader recognition, including the public, particularly if you make a significant discovery. Achievements in government are generally hidden from the public eye unless you have a very prominent government position—few do. Having said that, it must also be said that academia frequently values some time spent in government—experiences of use to the academic enterprise. And let's not forget that working for the collective good is a reward in itself for most of us.

Timescales are also different. A paper provides some sense of reward immediately after it is published and your name is on it. Work in government, such as a new policy or program, can take years to be identified as public good, and, as already stated, while understood internally, externally, your name may not even be specifically associated with that outcome.

Again, thinking about the individual versus the collective is also a good basis for really assessing your motives in considering one career path over the other—be honest with yourself.

## Rule 3: Government is more hierarchical than academia

With regard to hierarchy, in my opinion, a star researcher will quickly rise through the ranks and gain tenure. In government, it is more stepwise and experience related. Beyond promotion, reporting structures, while hierarchical in both environments, are adhered to more directly in government. In academia, the hierarchical relationship between a faculty member, department chair, and dean is there as an organizational structure but applied less rigorously than in most branches of government, notably the military.

Consider how much you like working in a structured environment when choosing between academia and government.

## Rule 4: Government offers better job security

Only a fraction of those on the academic ladder gain tenure (i.e., guaranteed salary) and, increasingly, in US parlance, they gain a partial full-time equivalent (FTE), which means that without supplementing your salary with grants, you can't survive. In other words, given the difficulty today in sustaining grant funding, your academic position is likely not very secure. Thus, depending on how attractive your field of research is to other employment sectors, you could be facing financial hardship. Government jobs often provide more security. Unless the government runs out of money or a department is closed, an unlikely but not unheard of event, you have a job. Thus, there are lots of government jobs that are essentially permanent. For shorter-term definable tasks, contractors are used by the government, for which the government has no obligation after the contract expires. Be sure to understand what type of government job you are applying for.

## Rule 5: Academia generally pays better

The downside of Rule 4 is that government jobs, at least in the US, typically do not pay as well as comparable positions in academia. While true as a general rule, as George Bernard Shaw once wrote, "The only golden rule is that there are no golden rules." So at the risk of following this statement immediately with a rule, academics tend to have a rather distorted view of what the government can actually pay its employees. This is surprising because you can typically find federal salaries online. The Freedom of Information Act in the US and similar legislation in other countries led to the creation of third-party websites that provide government salary information. This is easily compared with at least state-run academic institutions, which also make this information available. Private institutions are another matter. Explore the possibilities online.

## Rule 6: Both require persistence and patience but in different ways

Red tape plagues any organization of size. The bigger the organization, the more the red tape. Is it proportional to the size of the organization? Let me answer from my own experience. The US National Institutes of Health (NIH) budget is about $32,000,000,000 per year. The research budget of the University of California San Diego is approximately $1,000,000,000 per year. Is the NIH 32 times more bureaucratic? Probably not, but it is significantly more bureaucratic. Some of this serves a purpose. Consider an example. If the NIH makes a policy, it affects the whole of biomedical research in the US and likely beyond. If a Principal Investigator in academia makes a policy, it typically affects little more than that scientist's lab. NIH policies require legal scrutiny, a period of posting for public comment, and more. In other words, those advocating for the policy need persistence and patience to get it enacted. When that policy finally goes into effect, it has broad-ranging implications. Research obviously requires persistence and patience when, for example, an experiment is not working for unknown reasons, whereas in the case of a government process, the workflow is typically known. The time point to completion in government can be estimated; in academic research, it cannot. In government, the process is often out of the hands of the individual; it is less so in academia. While persistence and patience are required for both academia and a government career, the reasons for persistence and patience are different. I would suggest that it requires a different type of personality for each. Consider how your own persistence and patience match to each environment.

## Rule 7: It is harder to effect change in government, but changes are more likely to persist

As alluded to in Rule 6, an upside of working in government is that, when policies or other actions do get put into place, it is harder for them to be undone. It may require yet a new policy or action to replace the old, which takes time. There needs to be a good reason for the change, and thus, generally speaking, actions taken in government are persistent and hopefully the rewards long standing. Some folks gain satisfaction in knowing this and work well in government. Others are, well, too impatient, as per Rule 6.

Anyone who has sat through an academic faculty meeting might be tempted to say that it is harder to effect change in academia. I would still argue that faculty meeting outcomes generally take less time but have fewer ramifications relative to a government-made decision. Be prepared to work on a longer timescale in government to get your objectives accomplished.

## Rule 8: Working with the private sector is different

Academia and government treat working with the private sector differently, even though, in my experience, government workers are as innovative as their academic counterparts. While academia and government want the translation of research into products, the motivation is different. US academia has the Bayh-Dole Act, which actively encourages the commercialization of university research, and there is more of a direct financial incentive to the inventor and institution. Government workers are far less likely to profit directly from their innovations, and the government has less direct incentive to make money. Companies developing products from government innovations pay taxes, and so money gets fed back into government indirectly rather than directly, as is true of academia.

While perhaps not an innovation per se, government also provides less incentive to publish materials that return royalties, notably books. In the US government, the publisher retains royalties on books, whereas in academia, the author gets the royalties. In summary, if you want to be an entrepreneur, it will be easier in academia.

## Rule 9: Accountability is on a different scale

Government service is generally held to higher ethical standards than academia. This does not mean that government employees are more ethical than their academic counterparts. It simply means that the scale of possible malpractice is different, and the respective academic and government institutions respond to differing degrees. Moreover, an academic institution is responsible to a board of trustees; government is responsible to the public—significantly different levels of accountability. As a result, government responds to even the appearance of malpractice. Consider an example. Being on the scientific advisory board of a company while in academia is typically encouraged. It's good for the institution to have their faculty recognized in this way. The academic likely benefits from having shares in the company they are helping. Obviously, there are still ethical considerations for academics, such as applying grant monies, originally awarded for a different purpose, to benefit the company they are consulting for. By analogy, a government employee could influence the use of public monies to benefit a company and receive remuneration from that company. Both academic and government scenarios relate to an issue of trust. However, one is damaging to an individual and their academic organization, the other to all of government. Consequently, the ramifications are proportionally different. As a result, government employees are typically subject to tighter rules on what they may and may not do. So for example, if you want to be on scientific advisory boards or consult in a variety of other ways, academia is probably a better choice for you because this is typically not allowed as a government employee.

## Rule 10: Access to resources is different

In general, the government has more access to resources than academia, which are very much "soft money" institutions, where funding is unpredictable and of short duration. More than just a stable source of funding, government has real data, which can be attractive. Think of the National Security Agency (NSA). In the era of data science, characterized by the integration of disparate data sets, government can offer access to data not available to academia to conduct important studies relating to, for example, socioeconomic status and health.

Several years ago, when considering government service while being in academia, I made a pros and cons list to compare the 2 career paths. Hopefully, these rules will help you in creating such a list for yourself. Better still is the hope that others will comment on these rules to provide yet further insights.

Making the career choice presented here is daunting at any stage of one's career. Furthermore, perspectives may change in one's career; while academics is more desirable at one stage, service to a community may feel more rewarding at another. It is my hope that these rules will help in weighing trade-offs at any stage of one's career. Personally, I have thoroughly enjoyed my time in both academia and government, and I have no regrets in switching from academia to government and now back again. But then again, no regrets is my mantra for everything.

## Acknowledgments

Thanks to Daniel Mietchen for helping enrich this article with his own experiences.

## References

1. Searls DB (2009) Ten Simple Rules for Choosing between Industry and Academia. PLoS Comput Biol 5(6): e1000388. https://doi.org/10.1371/journal.pcbi.1000388 PMID: 19668326

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for getting the most out of a summer laboratory internship

**Toby P. Aicher[1‡], Dániel L. Barabási[2‡], Benjamin D. Harris[3‡]*, Ajay Nadig[4‡], Kaitlin L. Williams[5‡]**

**1** Department of Molecular Biology & Biochemistry, Middlebury College, Middlebury, Vermont, United States of America, **2** Department of Physics, University of Notre Dame, Notre Dame, Indiana, United States of America, **3** Department of Biology, Colgate University, Hamilton, New York, United States of America, **4** Department of Psychology, Northwestern University, Evanston, Illinois, United States of America, **5** Department of Life Sciences, Carroll University, Waukesha, Wisconsin, United States of America

‡ All authors contributed equally and should be considered co-first authors.
* bharris@colgate.edu

## Overview

Potential future scientists often gain their first exposure to real research practice through summer laboratory internships. Although these brief laboratory experiences are a major component of many public and private training initiatives, few written guide materials specifically address summer internships and how to optimally benefit from them. With that in mind, we have drawn on our summer research experiences to propose tips on how to approach all aspects of a summer internship, including planning ahead, navigating professional relationships, and maximizing impact, among others. We hope to open a conversation on how to enhance student experience at a critical early career juncture, when talented students are deciding whether to pursue a career in scientific research.





**Citation:** Aicher TP, Barabási DL, Harris BD, Nadig A, Williams KL (2017) Ten simple rules for getting the most out of a summer laboratory internship. PLoS Comput Biol 13(8): e1005606. https://doi.org/10.1371/journal.pcbi.1005606

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** August 17, 2017



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Perhaps you're working in a laboratory at your college or traveling to a new institution in a far flung location. In either case, a research internship can be an incredible stepping stone in your professional, intellectual, and personal development. Over the summer, you can learn cutting edge techniques, expand your network, and refine your interests as you look towards graduate school or the job market.

The process of obtaining a summer internship can be as long and subjective as applying for or choosing an undergraduate institution. Summer research programs are increasingly competitive—some such as Amgen programs have 1,000 to 2,000 applicants for just 2 dozen spots. The strategies and specifics of acquiring a summer research internship could be a "Ten simple rules" post of its own, so we point readers to articles and tools that have explored this topic previously (S1 Table). Keep in mind that you don't have to be at a prestigious institution to learn new techniques, gain exposure to a new research field, and connect with scientists and other scientists-in-training.

Guidelines have been presented in this article series on how to approach undergraduate research in general [1]. Building on these helpful rules, we observe that summer research internships present unique challenges due to their immersive and time-limited nature. That being said, these short-term positions present a unique set of challenges that may prevent interns from getting the most out of the summer months. For instance, research projects often operate on the timescale of years, leaving it unclear how to best spend an 8 to 10 week internship. With busy summer schedules and unclear expectations, interns may end up completing laboratory procedures without intellectually engaging in their project. To help interns and mentors navigate these and other related issues, we have compiled advice based on our collective 16 summer research experiences. We hope that these suggestions will help interns optimally learn from and contribute to their lab. More broadly, by sharing these tips, we hope more interns will experience the thrills of laboratory research that have led all of us to pursue careers in science.

## Rule 1: Plan ahead!

Any project that you work on as an intern is probably part of an effort that spans multiple years. To have a meaningful internship experience, you need to have a small slice of that project that is both significant and doable in a short amount of time. It is not easy or obvious for your mentors to design a summer project that achieves both of these goals.

Planning ahead with your mentor can help make sure the internship experience is productive and engaging. In general, mentors who take on undergraduate summer students want to create great projects, but busy schedules may hinder extensive planning. Postdocs and graduate students may learn that they are expected to advise a summer student with a week or two of notice, and must cobble together suitable projects.

As soon as you are accepted to a summer internship, reach out to your PI to express (1) your excitement at working in the lab, (2) that you are eager to begin discussing your project with your direct supervisor, and (3) ask for reading material to gain background for further correspondence. Repeat these sentiments prior to your arrival as well, as a reminder of your upcoming presence. As you work through preliminary readings, connect back with your supervisor with sections that appealed to or challenged you. If you are able to successfully communicate in this way, supervisors can make subtle decisions in experimental animal allocation, surgery timing, and reagent ordering to facilitate a great summer project.

## Rule 2: Own your project

Your internship, while exciting and maybe entirely novel for you, will be a blip on the timeline of your scientific career. To really gain competence, you have to put a considerable amount of time and effort into learning and mastering new skills. Take the extra 3 steps to really understand why you're pipetting liquid A into liquid B, and not the other way around. Recognize that protocols have intrinsic meaning, and you should know those meanings. You have been selected for what is likely a competitive and prestigious internship. Don't be a benchwork drone. What makes you more valuable than the thermocycler? What are you doing? Why are you doing it?

In addition, owning your project means asking for help when necessary. It is, however, easy to fall into the trap of seeking continuous and excessive oversight from your coworkers or mentor. You want to ask for assistance and goals, but not instruction for every small step included in your project. Asking once is okay. Asking twice simply because you didn't bother to write down the answer is annoying. If you have conducted research at your home institution, you will know that laboratory skills are often learned via trial and error over a long period of time. Summer internships afford no such luxury; it is critically important to optimize your learning by working diligently and efficiently. In cases where you continue to struggle with a concept, such as when your project has shifted directions midway, consider asking for a physical source, such as a paper or textbook, with which you can come to grips with it on your own.

It seems simple, but many undergraduates believe that they can remember everything said to them within the first few critical days in the lab. Your lab will likely gift you a notebook on your arrival, before they begin to enlighten you with the inner workings of every aspect of a protein you could imagine. Bring your lab notebook and writing instrument everywhere, and write down everything, thus developing a record of crucial methods, results, and insightful observations.

## Rule 3: Be humble

As an intern, you have to learn new techniques, get used to a new lab, and start a project that likely won't be finished by the time you leave. So, while you should strive to do the best research possible and generate data useful to the lab, you're unlikely to make a breakthrough discovery or get your name on a publication. Get used to this idea. Your mentor and your PI took you on because they think training the next generation of scientists is important and want to help you become a better scientist.

Be courteous to your mentor, PI, and lab. They made a choice to hire you as an intern, so make it worth their while. Always clean up your workspace when you're done, attend lab meetings, and arrive to lab when your mentor requests. Don't do anything inconsiderate to other members of the lab. Give your mentor and PI a thank you card or symbolic gift when you leave. You have a short time to make a good impression, and while your relationship with your summer PI may not be as well-established as with your home institution's PI, you can still form a positive, potentially long-enduring, connection.

Don't be too hard on yourself if you feel your project is moving slowly, isn't going anywhere, or isn't yielding interesting results. You should shoot for doing impressive work that advances a project but don't let that define your experience.

## Rule 4: Be a team player

Your internship project probably represents a small sliver of a much larger project. Understand your lab's mission and figure out how a summer intern can best contribute. This will change based on a number of parameters, including the stage of the project, the career stage of your supervisor, and how your lab utilizes the summer months.

Suppose your advising postdoc arrived in the laboratory only a few months before you. In this case, you should learn how to help set up equipment and build analysis pipelines that can be used long after you're gone. Alternatively, assume that your supervisor is in their sixth year with one foot out the door. Think of new analyses, and ask if you can play around with old data or help conceptualize results as they are being written up. If your lab supervisors are travelling to conferences all summer in a revolving door fashion, learn how to maintain their experiments and analyses, and become a reliable lab member who can hold down the fort. By recognizing what the situation is in your project and targeting your effort, you can have a larger impact on your lab and become an invaluable part of your team.

## Rule 5: Be a good collaborator

A successful internship is heavily dependent on having a good relationship with your mentor. Your mentor is responsible for giving you a project, teaching you new techniques, showing you around the lab, and helping you with problems that arise in your project. It is important that you establish a productive stream of communication with your mentor. Tell them if you don't like your project or if you want to change directions. Summer research projects tend to be more narrowly focused than long-term undergraduate research goals, and with good reason. If you think your project is vague and undefined, ask your mentor or PI if they can give you more concrete work with better defined goals.

The goal of a summer internship is not to be a lab technician or an extra pair of hands to assist your mentor. Act as a collaborator instead of as an assistant; work to develop the concepts and methods of the project, rather than robotically completing task lists provided to you. Once you have put in the legwork to understand your project, think critically about it: Can the experimental paradigm be modified to better answer the research question? Are there newer analysis strategies that can facilitate greater insight into your data? Respectfully engage your supervisor on these ideas. Although research design may seem complex and out of reach for an intern, you are probably closer to this level than you think, and summer projects provide a great opportunity to hone your inquiry skills in a low-stakes environment with a supervisor to guide you. In doing so, you will get a better feel for the full scope of independent research, leave a stronger impression on your lab, and get a better letter of recommendation.

## Rule 6: Meet your PI 90% of the way

In many labs, face time with the PI is a precious resource, and access for interns may be limited. This is especially the case in the summer months, which are frequently used for conference travel and writing. Realize that although this is the case, your PI took on a summer intern for a reason, and generally has the best intention of being an active mentor. However, this will not be presented to you on a silver platter, and you must seek it out.

Is your PI travelling for most of the summer? E-mail them, and ask if they have any availability in the next month for a lunch. Does your PI mostly run between meetings during the day, and work in their office only at night? Stick around the lab at night, and pop into their office with intelligent questions about your project or one of their papers. Does your PI generally only meet with postdocs? Ask your supervisor if you can sit in on a meeting to discuss issues related to the project you are working on. Strong connections with PIs for undergraduates in science are rarely a given; they are mostly earned by go-getting mentees who are willing to meet advisors 90% of the way.

## Rule 7: Get to know your whole lab

Coming into your internship you most likely have a defined image for the ongoing work of your lab based on your literature review, e-mails, interviews, and project description. Unless you find yourself in a particularly small or specialized group, this will only be part of the picture. Some of your newfound coworkers may be working on problems that seem unrelated to your project. Your lab will also have a diverse scientific background amassed from pulling researchers from various undergraduate and graduate institutions.

Learn a bit about everyone you work with, and get at least a cursory understanding of their projects. For a larger group this may prove complicated. Say your mentor is dedicated to microbiology, but another portion of the group is heavy on the analytical chemistry, and the rest are sealed off in the nanotech room. Wrapping your head around these topics at various stretches of academia will take a lot of time you don't have.

Instead, consider why individuals with diverging interests would be all under one group leader. Does nanotech provide some fundamental advantage to the analytical chemistry researchers, who can then aid the microbiologist aiming to halt the spread of disease? There must be some rhyme or reason, and understanding interdisciplinary interactions, large and small, will help you know where to look for new techniques later in your career.

It may be difficult during a few short weeks to become familiar with other researchers in your lab, especially if it is a large lab or people are away for summer conferences and presentations. Take advantage of going to lab meeting or going to lab social events to meet people in your lab and form connections. These gathering are a chance to talk about shared scientific interests and to connect to people on a personal level, both of which can help establish future professional connections.

## Rule 8: Get a feel for your data

Quantitative skills have become increasingly important in both research and industry career paths. Regardless of whether or not your internship has a programming course, you can start learning computational skills on your own. The simplest and easiest place to start learning quantitative skills is data analysis and presentation. Communicating information in posters, papers, or presentations requires you to interpret and display your data. In many cases you could get away with using software like Excel. But, by delving into more flexible and sophisticated tools, not only will you be able to have a more scientifically stimulating experience, you will also develop a skill set that is highly portable and will serve you, even if you don't continue in research. Instead of using Excel or SPSS statistical software to analyze and display your results, try learning a programming language such as R or Python. There are plenty of online tutorials and resources specifically targeted for data analysis, including the article "Ten simple rules for better figures" published in *PLOS Computational Biology* [2].

In addition to statistics and figures, you could explore the theoretical basis of your research with quantitative skills. Biological systems often have mathematical models that represent the system over time—neurons, cardiac cells, circadian rhythms, gene networks, and metabolic processes all can be represented using sets of differential equations. Learning how to implement your related model and create simulations in a programming language will not only build your understanding of your research topic, but also help discover new dynamics of the biological system. Computational projects are great side projects for a summer of research because you can pick one up and put it down whenever it is convenient since it does not cost extra laboratory resources. The computational skills you develop, regardless of how far you delve into them, will be useful for any career path you choose.

## Rule 9: Balance reading and doing

Understanding scientific literature is an integral part of the research process, and a summer research position provides an opportune time to strengthen your reading skills. Your summer will most likely begin with background articles from your mentor, and you can continue to explore implicated concepts in weekly in-lab journal clubs. Additionally, many programs have journal clubs specifically for interns. Join one and become an active member; discussing the contents of a paper is an excellent way to check your understanding. One author found that they digest the article best if they skim it on the computer, then print it out and annotate. With every paper, consider the experimental procedure, how the analyses was performed, and finally how these connect to the author's conclusions. If still in doubt, seek other articles that provide advice on reviewing scientific literature, but also seek your mentor's and peers' advice on how they tackle their reading.

However, you're only at your internship for 10 weeks, and reading literature takes time away from working on your project. Striking a balance will help you get the most out of your time. This all depends on your background and your project. You can get by with reading enough to grasp an understanding of your project. For some well-defined projects that may be only a couple of papers. For projects that require you to design or select methods that may be a lot more. If your summer experience is in the same field as your home institution research, focus more upon specifics of your project and lab. Many internship programs assign students to their labs and projects, so even if you preference something you have a lot of experience in, you could end up in a totally different field. If you are entering an entirely new territory—like a cancer researcher entering a plant laboratory—spend some additional time beefing up your general knowledge of the field.

At some point you will be put on the spot to get your opinion on someone else's work in the lab or on a paper the whole lab read. "I don't know" is always an acceptable answer, but it shouldn't be your default.

### Rule 10: Have fun

If all the above points on this list pan out, you're going to have a thrilling experience, complete with a satisfying project, a supportive group, and unique social interactions. Your workday may begin with breakfast with coworkers, continue with in-depth research and group meetings, which will leave you tempted to follow-up with journal reviews in the evening. Immersing yourself like this will be exhausting!

It'll prove beneficial to force yourself to step away from science every so often. The authors fondly remember playing volleyball as a break, thereby getting the exercise your brain and body needs. Take weekends to explore around your workplace. If you're in a new location, then you have a myriad of novel opportunities to uncover nearby, but even if you stayed at your home institution, this could be a chance to move outside the collegiate bubble. Your best inspiration may come from an art gallery or a movie; giving your conscious scientific focus a break will let your subconscious make groundbreaking connections.

Depending on your excitement and work ethic regarding your project, this may be uncomfortable, even difficult. If you're feeling guilty about taking some time to yourself, consider this: you may be the only one not doing so. At the end of the day, your PI and mentors go home to their families or roommates. They have dinner, maybe watch a movie with their kids. On weekends, they go on hikes, to a cool new art gallery, or a performance downtown. You probably do the same at college: classes, research and other commitments may take up most of your week, but the time in between is filled in with relaxing on the couch with friends. Research is a marathon, not a sprint, so take a break to leave yourself the stamina to stay in the race!

## Conclusion

Through the "Ten simple rules" and corresponding narratives, we explored the difficulties and realizations we have experienced through our summer internships. By addressing these hurdles, we hope to provide a smoother experience for you, and help you accomplish more with bolstered confidence. As final takeaway, we should note that your internship doesn't end with your last day. Your coworkers, from peers to mentor to PI, will be interested in your ongoing trajectory. Keep in touch about your accomplishments, and consider any and all connections you have made as your greatest career resources. Your mentors should be happy to point out ongoing opportunities in the field, and perhaps more open than others about what to avoid on your scientific journey. Cherish this support, and be sure to pay it forward with your own mentees as you advance in your career.

## Supporting information

**S1 Table. Online resources for finding a summer internship.**
(DOCX)

## Acknowledgments

We would like to thank all of our mentors and Principal Investigators from over the years for their insightful comments on this manuscript and their support in science.

## References

1. Yu M, Kuo YM. Ten simple rules to make the most out of your undergraduate research career. PLoS Comput Biol. 2017 May 4; 13(5):e1005484. https://doi.org/10.1371/journal.pcbi.1005484 PMID: 28472033
2. Rougier NP, Droettboom M, Bourne PE. Ten simple rules for better figures. PLoS Comput Biol. 2014 Sep 11; 10(9):e1003833. https://doi.org/10.1371/journal.pcbi.1003833 PMID: 25210732

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Developing a Successful Research Proposal in Brazil

**Dyoni M. de Oliveira[1], Marcos S. Buckeridge[2], Wanderley D. dos Santos[1]***

**1** Laboratory of Plant Biochemistry, Department of Biochemistry, State University of Maringá, Maringá, Paraná, Brazil, **2** Department of Botany, Institute of Bioscience, University of São Paulo, São Paulo, São Paulo, Brazil

* wdsantos@uem.br

## Introduction

Writing well is fundamental to publishing and having a successful scientific career [1], and being able to write a good research proposal is critical for obtaining financial support [2]. In emerging economies, such as Brazil, it is necessary to confront drawbacks not encountered in high-income countries [3]. The developing world has growing investments in science, technology, and innovation in many areas [4–6], including computational biology [7]. These investments have produced positive results in scientific quality in developing countries [8]. Although this is remarkably positive, the emergence of high-level research groups creates a highly competitive environment. We suggest a roadmap of ten simple rules for writing a consistent and convincing research project, which may be useful for researchers in Brazil and other emerging economies. There are several funding agencies in Brazil, and two of them—the National Council for Research Development (CNPq) and the São Paulo Research Foundation (FAPESP)—are used as examples of how proposals can be better adjusted in order to be successful. The latter represents the state funding agencies. Our ten rules will consider these agencies as the generic targets of proposals. When describing the ten rules below, we consider applications for research grant proposals and for MSc and PhD fellowships.



OPEN ACCESS

**Citation:** de Oliveira DM, Buckeridge MS, dos Santos WD (2017) Ten Simple Rules for Developing a Successful Research Proposal in Brazil. PLoS Comput Biol 13(2): e1005289. doi:10.1371/journal.pcbi.1005289

**Published:** February 2, 2017



**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Rule 1: Define the Problem Clearly

In general, the most important part of a research project is to precisely define the problem to be investigated. If you wish to ask for financial support for your research, it is imperative to attest that your interest is in line with research supported by the funding calls available. Some calls are generic and flexible, such as the Universal Call from the CNPq in Brazil. It is not essential to fit into specific calls, but it is certainly easier to swim downstream whenever possible. The relevance and originality of your research targets and, of course, internal coherence (between targets and methods) have a significant impact on the value of the project. An extensive and updated review of the relevant literature can guarantee the originality of your targets, averting as much as possible the risk of your findings being published by another group while your project is still ongoing. In this case, the risk is producing good results without relevance to your field. Although the originality and relevance of a proposal can be ensured, the relevance of your findings is unpredictable.

Originality is usually inversely proportional to risk. When an idea is proposed, the level of novelty may lead a reviewer to find the project too risky. When preparing your proposal, it is therefore important to describe the risks very clearly. Brazilian reviewers tend to be quite conservative, and even a low risk level can be considered too much risk. When reviewers of FAPESP, for example, are completing the review form, there are boxes at the end that have to be ticked. If a box like "Very Good with Minor Deficiencies" is ticked, the coordinator (the level above a reviewer) who will make the final decision may hesitate to approve the proposal. Authors thus have to be very careful and clearly explain the risks related to the project in order to minimize the possibility of the reviewer ticking the boxes that point out deficiencies. Because this is a cultural problem related to the reviewers rather than to the applicants, it is very important that the applicant display preliminary results that clearly and elegantly show the reviewer that the risks are manageable. This is difficult and demands hard and careful thinking, but it is the only way to change the conservative culture of Brazilians into one that incorporates a more open and braver view of the work in science.

## Rule 2: Formulate Falsifiable Hypotheses and Include Preliminary Data

Sometimes, you can summarize your research as a precise and complete survey of data; however, when studying complex systems (such as living beings), measuring everything might not be feasible or convenient. It can therefore be useful to formulate a hypothesis that you can test with a number of experiments. You must formulate the hypothesis as an affirmative, clear, and concise sentence (e.g., "The volume of the liquid water is directly proportional to the temperature"). This statement must express an up-to-date possibility based on a systematic review of scientific knowledge on the defined theme; however, you cannot know beforehand whether your hypothesis is really correct (as in the example above, which we currently know to be wrong). The important thing here is to ensure that your hypothesis is testable under the actual conditions you have or have access to (physical, financial, and human resources) so that you can develop plausible experiments to test it. In low-income countries, being creative in the proposition of accessible methodologies for testing a hypothesis is especially critical, as we discuss in Rule 5.

Often, the hypothesis of a research project arises as a result of previous observations, and presenting the preliminary data might provide crucial support for your hypothesis. The preliminary data will also help you to effectively convince a reviewer that you have the technical and scientific expertise to carry out the work as proposed[2].

Again, try to prepare the text so that the reviewer concludes, after reading the project, that what you want to do is indeed science and that it is more than that: it is good science that advances knowledge.

## Rule 3: Establish Clear Objectives

After formulating your hypothesis (or hypotheses), you should establish a clear and explicit goal, the necessity of which is exemplified in the excerpt below:

> *Finding herself lost in Wonderland, Alice asked to the Cheshire Cat: "Would you tell me, please, which way I ought to go from here?" "That depends a good deal on where you want to get to," the cat replied. "I don't much care where—" said Alice. "Then it doesn't much matter which way you go," said the cat* [9].

You can find much by chance or serendipity; however, what you discover by chance will not necessarily be the same thing you were looking for. In a project, you want to convince others to trust your goals and that you are competent. A clear goal will guide the choice of the methodologies that you will use to get there. Objectives underlie an experimental design and can serve as a basis for performance evaluation and a change of strategy, when necessary. During the execution of the research, you will probably have to divide your attention across many tasks: classes, paperwork, other projects, supervisions, and so on. A clear statement of the objectives will remind you and your collaborators of where are you going and how you intend to get there.

Most reviewers are busy scientists, and they have to perform a great deal of administration along with their scientific work. In some cases, they will read the objectives (and the title and abstract) more carefully than other parts of the project. Therefore, be absolutely sure that you are describing your objectives in a simple way.

## Rule 4: Estimate the Duration and Requirements of Experimental Procedures Carefully

If your goals define the specific procedures you will follow during the research, the reverse is also true. The design of your experimental approach will also help you define the goals of your proposal. To reach specific goals, it is important to gain access to the necessary in-house facilities or invite external collaborators, as previously discussed in Ten Simple Rules [10]. If you have no expertise, this can be crucial. Talk to experienced researchers about the techniques you have in mind. Ask how long procedures specifically require, and be careful about laconic answers—especially from your mentor! They can reflect optimistic expectations or the desire to obtain results without thinking deeply about realistic deadlines. Be sure that the time required to carry out experimental procedures is compatible with the maximum period established by the funding call. It is important to know as much as possible about experimental methodologies in order to avoid a design that is impracticable within the timeframe and with the resources available. Calculate the required time, allotting sufficient time for replanning. Bear in mind that scientific research is full of unpredictable mishaps (but also serendipity), and thus, it is important to evaluate the possible risks of things that do not go well. By identifying these risks, you can attempt to avoid them when managing your research. The funding agencies and especially private funders expect you to fulfil what you promised in the proposal, even after the deadline (and in this case, without additional resources!), and thus, it is always wise to promise the minimum necessary to achieve your goals. In Brazilian science, this is the most important failing point in proposals. Reviewers are usually not aware of those failures, however. Proposals in Brazil rarely include a schedule showing clearly when each milestone of the project should be reached, by whom it will be produced, and how the different tasks are associated with the objective of the project. However, this is one of the most important parts of the project because it gives the reviewer a clearer idea about the feasibility of the proposal. Your project may be original, the objectives may be clear, and the methodology choices may be appropriate, but if you do not construct a framework of tasks and resources (people and money) that are clearly coordinated, the reviewer will not be able to evaluate the feasibility of your project. Brazilian funding agencies usually fund a relatively small percentage of the proposals submitted. Final decisions are also made in a comparative review, in which a board of reviewers may decide together who will be approved. If your competitors have a more detailed schedule, they will thus have an advantage, and your proposal is more likely to be turned down.

## Rule 5: Explain the Methodologies for the Goal, to Demonstrate That You Can Carry out the Research

Provide methodological descriptions that best fit your needs, your knowledge, and your financial reality. Take special care not to write methodologies that are incomplete or inapplicable to your particular case. It is common to find inconsistencies in proposals due to the "copying and pasting" of methodologies from other proposals. A zealous reviewer may require correction, and you may find yourself in a difficult situation or be asked to correct your work, particularly if you have to defend your proposal in public. Remember that there are two kinds of knowledge: tacit and explicit. Written methodologies usually hide important details that belong to the domain of tacit knowledge. You learn the tricks of the trade only by practicing and training with an experienced researcher. If you only have access to explicit knowledge to perform an experiment, you will probably make mistakes. The person evaluating your proposal, who is usually specialized in the field, might consider this restriction by consulting your curriculum vitae. You can lessen the potentially negative impact of this problem through careful planning, which allows a surplus of time for establishing a protocol [11].

Some scientists and reviewers think that the methodology is the most important part of a proposal, so be certain that you are using (1) the right methods for the purpose of each experiment and (2) a currently accepted methodology. This does not always mean that you should only use the most advanced technology. Reviewers usually base their evaluation on the basis of a trade-off between the novelty of the method and the adequacy of using it, especially in light of how much money you are going to spend to perform the experiments.

## Rule 6: Clearly Define the Tasks, People in Charge, and Costs in Your Research Proposal

In order to answer a scientific question, it may be necessary to complete a series of goals and perform a series of experiments. It is thus useful to clearly define the following points for each goal: (1) What are the dates for initiating and finishing the experiment? (2) Who will carry out the experiment (in the case of a group)? (3) How much will the experiment cost? (4) How will you assess the research progress? (5) What are the critical risks? (6) How might you deal with severe problems? Even with good maintenance, equipment might fail, or a technician become unavailable. Try not to underestimate the deadline required for a crucial task, and, if possible, identify spare facilities/specialists to whom you can resort, if necessary.

## Rule 7: Preventing the Unpredictable: Establish a Flexible Schedule

The result of one goal may be essential for the start of another. It is thus advisable for you to outline the best order in which each task must be executed. Nonetheless, remember that the schedule works as a possible way to execute all necessary tasks in your proposal. Not everything you plan must happen exactly as you originally conceived it, and it is natural for the schedule to undergo changes throughout the project development. As such, a good schedule must be flexible enough to accommodate the unpredictable obstacles you will face.

If you have a well-constructed schedule, this could be where the reviewer will look very carefully to find inconsistencies and point them out as a deficiency. If the percentage of approval is low, the probability is that reviewers will start reading your proposal by looking for deficiencies. When they find something, they will consider this as an inconsistency in contrast with a top-level ideal proposal. You should thus expect to be penalized for every small mistake found in your proposal, and a complex schedule is somewhere a reviewer may find many problems, thus turning down your proposal.

## Rule 8: Justify the Benefits Your Research Will Provide

An exhaustive survey of the relevant literature will help you introduce the field and convince reviewers why the problem you chose deserves attention. If you are sure your research program is unique and relevant, prove it to the experts who will judge your proposal, presenting a complete state-of-the-art picture. Be parsimonious, however, with words, and do not lose focus. Before writing, establish a briefing with the necessary information. Organize the ideas as an inverted pyramid, from general information about the field to the specific area that your work will address. This will make it easy for the reader to follow your reasoning and to understand the focus of your research.

After the introductory context, it is important to stress to the reader why your work is important. Emphasize the practical advantages (technological) that may result from your research and the importance that these results may have in overcoming the knowledge gaps mentioned in the introduction. This is the time to provide a convincing support for the relevance and importance of your research project. A list of scenarios resulting from your research can facilitate appreciation by the reader of the possible impact of your proposal as a whole.

Currently, there is a clear trend for applications to note the societal implications of the research proposed, so it is quite important to explain the main connections between the results you will produce and the benefits they will bring to society. This issue is more important to the higher-level members of the funding agency—scientists who design and maintain the general policies of science for the country—than to the reviewer. If your proposal is considered excellent in all the items above, but your explanation of the benefits to the country or region is not clear, your proposal might be turned down if competition is tight.

## Rule 9: Write a Good Title and Abstract

The title must inform the reader about the scope of the project. It is common in scientific literature to have the title briefly describe the issue addressed in the work (article or project). Avoid the use of adverbs and scientific nomenclature, which are not strictly necessary. Remember that scientists are attracted by intelligence. A creative title can arouse empathy in the reviewer.

An abstract is optional; however, it facilitates reading and understanding the general idea of the proposal. Although you must avoid prolixity at all costs, depending on the complexity of the theme, the introduction may become long and complex. This may mislead the reader if you do not properly formalize the focus of the work in an abstract. Together, the title and the abstract are good opportunities to put forward your idea. Most reviewers will make their initial decisions about whether to approve a proposal or not immediately after carefully reading the title and abstract, so do not underestimate them.

Again, many busy reviewers will read your title and abstract more carefully. Because scientists are quite busy, there is a tendency to use fast thinking rather than a slow and thoughtful analysis of projects [12]. Usually, after reading the title and the abstract, the reviewer will have already made a decision about whether his or her thumb will be up or down for the proposal. If your title and abstract are well designed, the reviewer will continue reading and will give you several other opportunities to sell your work. Check your title and abstract many times if necessary, and never leave spelling mistakes in them. Spelling mistakes and/or inadequate language at the beginning score much higher in the negativity scale of the reviewer than such mistakes in the middle of the text.

## Rule 10: Organize a Logical Structure and Make the Text More Readable

Your proposal should be concise and impart as much information as possible in the least number of words. After ensuring that your research has precise and feasible objectives, is well contextualized and justified, has a consistent schedule, and is convincingly introduced, it is time to review the text. In addition to the items outlined here (Title, Abstract, Introduction, and so on), you can add other items you find appropriate. Check for a model provided by the funding institution. If no such model is available, take a careful look at successful projects. Try not to be too creative in the way you organize your proposal. You do not need to be a copy machine, but try to respect practices already consolidated. Organize the text in order to make it enjoyable, educational, and accurate [11].

Review your text to find small mistakes that are easy to spot. An excess of easily detectable mistakes suggests laziness. Be careful with the bibliography, which is tedious to organize, because it is very easy to leave mistakes there. The accuracy of references is extremely important because a reader (reviewer) may at any time become curious and check one of them. Give preference to software that automatically organizes references, but also remember that you require a good word processing software program.

Language and semiotics have to be carefully adjusted by many people in the world, and Brazilians are not an exception. Brazilians do not like informal language, so the use of the words "I" or "we" in the text should be avoided. Although Brazilians are not usually direct when speaking, Portuguese has to be transformed into the English style for science texts, using short phrases and sparse punctuation. Discrete humility is important in the text. It is important to find the right balance regarding how you value your work.

## Acknowledgments

The style for this article was inspired by the "Ten Simple Rules. . ." papers published by Philip E. Bourne in *PLOS Computational Biology*. This work was inspired by the discipline BIB0306 offered to the undergraduate students of the Department of Botany of the Institute of Biosciences of the University of São Paulo.

## References

1. Zhang W (2014) Ten simple rules for writing research papers. PLoS Comput Biol 10(1): e1003453. doi: 10.1371/journal.pcbi.1003453 PMID: 24499936
2. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comput Biol 2(2): e12. doi: 10.1371/journal.pcbi.0020012 PMID: 16501664
3. Moreno E, Gutiérrez J- M (2008) Ten simple rules for aspiring scientists in a low-income country. PLoS Comput Biol 4(5): e1000024.
4. Noorden RV (2014) The impact gap: South America by the numbers. Nature 510(7504): 202–203. doi: 10.1038/510202a PMID: 24919906
5. Catanzaro M, Miranda G, Palmer L, Bajak A (2014) South American science: big players. Nature 510 (12): 204–206.
6. Regalado A (2010). Science in Brazil. Brazilian science: riding a gusher. Science 330(6009): 1306–1312. doi: 10.1126/science.330.6009.1306 PMID: 21127226
7. Neshich G (2007) Computational Biology in Brazil. PLoS Comput Biol 3(10): e185.
8. Thomaz SM, Mormul RP (2014) Misinterpretation of 'slow science' and 'academic productivism' may obstruct science in developing countries. Braz J Biol 74(3): S1–S2.
9. Dodgson CL (1865). Alice's Adventures in Wonderland. New York, D. Appleton and Co., p. 227.
10. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3(3): e44. doi: 10.1371/journal.pcbi.0030044 PMID: 17397252
11. Davies MB (2007) Doing a successful research project. Using qualitative and quantitative methods. London, Palgrave Macmillan, p. 25–26.
12. Kahneman D. (2011) Thinking, fast and slow. Farrar, Straus and Giroux, 499p.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Writing a Postdoctoral Fellowship

**Ke Yuan[1], Lei Cai[2], Siu Ping Ngok[3], Li Ma[4]*, Crystal M. Botham[5]***

**1** Division of Pulmonary and Critical Care Medicine, Department of Medicine, Stanford University, Stanford, California, United States of America, **2** Department of Materials Science and Engineering, Stanford Neuroscience Institute, Stanford University, Stanford, California, United States of America, **3** Division of Hematology, Oncology, Stem Cell Transplantation, and Cancer Biology, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, United States of America, **4** Asian Liver Center and Department of Surgery, Stanford University, Stanford, California, United States of America, **5** Stanford Biosciences Grant Writing Academy, Stanford University, Stanford, California, United States of America

* lma3@stanford.edu (LM); cbotham@stanford.edu (CMB)



**Citation:** Yuan K, Cai L, Ngok SP, Ma L, Botham CM (2016) Ten Simple Rules for Writing a Postdoctoral Fellowship. PLoS Comput Biol 12(7): e1004934. doi:10.1371/journal.pcbi.1004934

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** July 14, 2016



**Funding:** Dr. Ke Yuan is supported by American Heart Association Scientist Development Grant (15SDG25710448) and the Pulmonary Hypertension Association Proof of Concept Award (SPO121940). Dr. Lei Cai is supported by Stanford Neuroscience Institute and NIH NRSA postdoctoral fellowship (1F32HL128094). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Postdoctoral fellowships support research, and frequently career development training, to enhance your potential to becoming a productive, independent investigator. Securing a fellowship sends a strong signal that you are capable of conducting fundable research and will likely lead to successes with larger grants. Writing a fellowship will also increase your productivity and impact because you will learn and refine skills necessary to articulate your research priorities. However, competition is fierce and your fellowship application needs to stand out among your peers as realistic, coherent, and compelling. Also, reviewers, a committee of experts and sometimes non-experts, will scrutinize your application, so anything less than polished may be quickly eliminated. We have drawn below ten tips from our experiences in securing postdoctoral fellowships to help as you successfully tackle your proposal.

## Rule 1: Start Early and Gather Critical Information

Crafting a competitive fellowship can take 6–9 months, so it is imperative that you start early. You may even want to start looking for postdoctoral fellowships before you finish your doctoral degree. Compile a comprehensive list of fellowships that you can apply to. This list should include key information to organize your game plan for applying, including Sponsor (agency sponsoring the fellowship) name; URL for funding information; Sponsor deadlines; and any other requirements or critical information.

To find suitable fellowships, start by asking your faculty mentor(s), laboratory colleagues, and recent alumni about their experiences applying for fellowships. Federal agencies in the United States, such as the National Institute of Health (NIH) and National Science Foundation (NSF); foreign governmental agencies; and other organizations, such as societies, foundations, and associations, often solicit fellowship applications. Additionally, many institutions offer internally supported fellowships as well as institutional research training grants.

Once you have an exhaustive list of fellowships you are eligible for, start gathering critical information that you can use to inform your writing. Read the fellowship instructions completely and identify the review criteria. Investigate the review process; NIH's Center for Scientific Review reviews grant applications for scientific merit and has a worthwhile video about the Peer Review Process [1]. Sometimes Sponsors offer notification alerts about upcoming funding opportunities, deadlines, and updated policies, so make sure to sign up for those when offered. Also, gather previously submitted applications and reviewers' comments for the fellowships you will to apply to. Both funded and unfunded applications are useful. Sometimes Sponsors make available funded abstracts like NIH's Research Portfolio Online Reporting Tools (RePORT), and these provide critical information about the scope of funded projects.

Many institutions have internal policies and processes that are required before a proposal can be submitted to a Sponsor. These requirements can include waivers to assess eligibility and internal deadlines (five business day internal deadlines are standard), so make sure you also gather relevant information about any internal policies and processes required by your institution.

## Rule 2: Create a Game Plan and Write Regularly

Writing a compelling fellowship takes time, a lot of time, which is challenging to balance with a hectic laboratory schedule, other responsibilities, and family obligations. To reduce stress, divide the fellowship requirements into smaller tasks by creating a detailed timeline with goals or milestones. Having a game plan with daily and/or weekly goals will also help you avoid procrastination. Make sure you are writing regularly (i.e., daily or every other day) to establish an effective writing practice. This will increase your productivity and reduce your anxiety because writing will become a habit. It is also important to make your writing time non-negotiable so other obligations or distractions don't impede your progress.

## Rule 3: Find Your Research Niche

It is crucial that you have a deep awareness of your field so you can identify critical knowledge gaps that will significantly move your field forward when filled. Keep a list of questions or problems inherent to your field and update this list after reading germane peer-reviewed and review articles or attending seminars and conferences. Narrow down and focus your list through discussions with your mentor(s), key researchers in your field, and colleagues. Because compelling projects often combine two seemingly unrelated threads of work to challenge and shift the current research or clinical practice paradigms, it is important to have a broad familiarity with the wider scientific community as well. Seek opportunities to attend seminars on diverse topics, speak with experts, and read broadly the scientific literature. Relentlessly contemplate how concepts and approaches in the wider scientific community could be extended to address critical knowledge gaps in your field. Furthermore, develop a few of your research questions by crafting hypotheses supported by the literature and/or preliminary data. Again, share your ideas with others, i.e., mentor(s), other scientists, and colleagues, to gauge interest in the significance and innovation of the proposed ideas. Remember, because your focus is on writing a compelling fellowship, make sure your research questions are also relevant and appropriate for the missions of the sponsoring agencies.

## Rule 4: Use Your Specific Aims Document as Your Roadmap

A perfectly crafted Specific Aims document, usually a one-page description of your plan during the project period, is crucial for a compelling fellowship because your reviewers will read it! In fact, it is very likely your Specific Aims will be the first document your reviewers will read, so it is vital to fully engage the reviewers' interest and desire to keep reading. The Specific Aims document must concisely answer the following questions:

- *Is the research question important?* Compelling proposals often tackle a particular gap in the knowledge base that, when addressed, significantly advance the field.

- *What is the overall goal?* The overall goal defines the purpose of the proposal and must be attainable regardless of how the hypothesis tests.
- *What specifically will be done?* Attract the reviewers' interest using attention-getting headlines. Describe your working hypothesis and your approach to objectively test the hypothesis.
- *What are the expected outcomes and impact?* Describe what the reviewers can expect after the proposal is completed in terms of advancement to the field.

A draft of your Specific Aims document is ideal for eliciting feedback from your mentor(s) and colleagues because evaluating a one-page document is not an enormous time investment on part of the person giving you feedback. Plus, you don't want to invest time writing a full proposal without knowing the proposal's conceptual framework is compelling. When you are ready to write the research plan, your Specific Aims document then provides a useful roadmap.

As you are writing (and rewriting) your Specific Aims document, it is essential to integrate the Sponsor's goals for that fellowship funding opportunity. Often goals for a fellowship application include increasing the awardee's potential for becoming an independent investigator, in which case an appropriate expected outcome might be that you mature into an independent investigator.

We recommend reading The Grant Application Writer's Workbook (www.grantcentral.com) [2] because it has two helpful chapters on how to write a persuasive Specific Aims document, as well as other instructive chapters. Although a little formulaic, the Workbook's approach ensures the conceptual framework of your Specific Aims document is solid. We also advise reading a diverse repertoire of Specific Aims documents to unearth your own style for this document.

## Rule 5: Build a First-Rate Team of Mentors

Fellowship applications often support mentored training experiences; therefore, a strong mentoring team is essential. Remember, reviewers often evaluate the qualifications and appropriateness of your mentoring team. The leader of your mentoring team should have a track record of mentoring individuals at similar stages as your own as well as research qualifications appropriate for your interests. Reviewers will also often consider if your mentor can adequately support the proposed research and training because fellowship applications don't always provide sufficient funds. It is also useful to propose a co-mentor who complements your mentor's qualifications and experiences. You should also seek out other mentors at your institution and elsewhere to guide and support your training. These mentors could form an advisory committee, which is required for some funding opportunities, to assist in your training and monitor your progress. In summary, a first-rate mentoring team will reflect the various features of your fellowship, including mentors who augment your research training by enhancing your technical skills as well as mentors who support your professional development and career planning.

As you develop your fellowship proposal, meet regularly with your mentors to elicit feedback on your ideas and drafts. Your mentors should provide feedback on several iterations of your Specific Aims document and contribute to strengthening it. Recruit mentors to your team who will also invest in reading and providing feedback on your entire fellowship as an internal review before the fellowship's due date.

You also want to maintain and cultivate relationships with prior mentors, advisors, or colleagues because fellowships often require three to five letters of reference. A weak or poorly written letter will negatively affect your proposal's fundability, so make sure your referees will write a strong letter of recommendation and highlight your specific capabilities.

## Rule 6: Develop a Complete Career Development Training Plan

Most fellowships support applicants engaged in training to enhance their development into a productive independent researcher. Training often includes both mentored activities, e.g., regular meetings with your mentor(s), as well as professional activities, e.g., courses and seminars. It is important that you describe a complete training plan and justify the need for each training activity based on your background and career goals.

When developing this plan, it is helpful to think deeply about your training needs. What skills or experiences are missing from your background but needed for your next career stage? Try to identify three to five training goals for your fellowship and organize your plan with these goals in mind. Below are sample activities:

- Regular (weekly) one-on-one meetings with mentor(s)
- Biannual meeting with advisory committee
- Externship (few weeks to a few months) in a collaborator's laboratory to learn a specific technique or approach
- Courses (include course # and timeline) to study specific topics or methods
- Seminars focused on specific research areas
- Conferences to disseminate your research and initiate collaborations
- Teaching or mentoring
- Grant writing, scientific writing, and oral presentation courses or seminars
- Opportunities for gaining leadership roles
- Laboratory management seminars or experiences

## Rule 7: STOP! Get Feedback

Feedback is critical to developing a first-class proposal. You need a wide audience providing feedback because your reviewers will likely come from diverse backgrounds as well. Be proactive in asking for feedback from your mentor, colleagues, and peers. Even non-scientists can provide critical advice about the clarity of your writing. When eliciting feedback, inform your reviewer of your specific needs, i.e., you desire broader feedback on overall concepts and feasibility or want advice on grammar and spelling. You may also consider hiring a professional editing and proofreading service to polish your writing.

Some fellowships have program staff, such as the NIH Program Officers, who can advise prospective applicants. These individuals can provide essential information and feedback about the programmatic relevance of your proposal to the Sponsor's goals for that specific fellowship application. Approaching a Program Officer can be daunting, but reading the article "What to Say—and Not Say—to Program Officers" can help ease your anxiety [3].

## Rule 8: Tell a Consistent and Cohesive Story

Fellowship applications are often composed of numerous documents or sections. Therefore, it is important that all your documents tell a consistent and cohesive story. For example, you might state your long term goal in the Specific Aims document and personal statement of your biosketch, then elaborate on your long term goal in a career goals document, so each of these documents must tell a consistent story. Similarly, your research must be described consistently in your abstract, Specific Aims, and research strategy documents. It is important to allow at least one to two weeks of time after composing the entire application to review and scrutinize the story you tell to ensure it is consistent and cohesive.

## Rule 9: Follow Specific Requirements and Proofread for Errors and Readability

Each fellowship application has specific formats and page requirements that must be strictly followed. Keep these instructions and the review criteria close at hand when writing and revising. Applications that do not conform to required formatting and other requirements might be administratively rejected before the review process, so meticulously follow all requirements and guidelines.

Proofread your almost final documents for errors and readability. Errors can be confusing to reviewers. Also, if the documents have many misspellings or grammar errors, your reviewers will question your ability to complete the proposed experiments with precision and accuracy. Remove or reduce any field-specific jargon or acronyms. Review the layout of your pages and make sure each figure or table is readable and well placed. Use instructive headings and figure titles that inform the reviewers of the significance of the next paragraph(s) or results. Use bolding or italics to stress key statements or ideas. Your final documents must be easy to read, but also pleasing, so your reviewers remain engaged.

## Rule 10: Recycle and Resubmit

Fellowships applications frequently have similar requirements, so it is fairly easy to recycle your application or submit it to several different funding opportunities. This can significantly increase your odds for success, especially if you are able to improve your application with each submission by tackling reviewers' comments from a prior submission. However, some Sponsors limit concurrent applications to different funding opportunities, so read the instructions carefully.

Fellowship funding rates vary but, sadly, excellent fellowships may go unfunded. Although this rejection stings, resubmitted applications generally have a better success rate than original applications, so it is often worth resubmitting. However, resubmitting an application requires careful consideration of the reviewers' comments and suggestions. If available, speak to your Program Officers because he or she may have listened to the reviewers' discussion and can provide a unique prospective or crucial information not included in the reviewers' written comments. Resubmitted fellowships are many times allowed an additional one- to two-page document to describe how you addressed the reviewers' comments in the revised application, and this document needs to be clear and persuasive.

## Conclusion

The ten tips we provide here will improve your chances of securing a fellowship and can be applied to other funding opportunity announcements like career development awards (i.e., NIH K Awards). Regardless of funding outcomes, writing a fellowship is an important career development activity because you will learn and refine skills that will enhance your training.

## References


1. National Institutes of Health. NIH Peer Review Reveal—a front-row seat to a review peer review meeting. https://www.youtube.com/watch?v=fBDxI6l4dOA.

2. Stephen W. Russell and David C. Morrison. The Grant Application Writer's Workbook–National Institutes of Health Version. Available: www.grantcentral.com.

3. Spires MJ. What to Say—and Not Say—to Program Officers. The Chronicles of Higher Education. 2012. Available: http://chronicle.com/article/What-to-Say-and-Not-Say-to/131282.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules to Win a Nobel Prize

**Richard J. Roberts***

New England Biolabs, Ipswich, Massachusetts, United States of America

* roberts@neb.com

## Preface

by Philip E. Bourne, National Institutes of Health, Founding Editor-in-Chief of *PLOS Computational Biology*

When receiving a draft of the article "Ten Simple Rules for Writing a PLOS Ten Simple Rules Article" [1], not only had we come full circle in terms of professional development, but also I knew the series was a success. Since that article was published in October 2014, two more articles have been published, and this will be the third: a total of 44 in all. Rule 2 in what I shall affectionately call the 102 article [1] suggested you need a novel topic and suggested winning a Nobel Prize was such a topic. As I hinted in my editorial comments to the 102 article, I would take up the challenge in soliciting such an article. Rich Roberts was the first person to come to mind, partly because he is a good sport, partly because we share an interest in open (to be interpreted here as candid) science, and of course because he won the Nobel Prize in Physiology or Medicine (with Phillip Sharp) in 1993 for work on gene structure.

At first he was reluctant and slightly insulted, making me think I should write "Ten Simple Rules for How Not to Insult a Nobel Laureate." The rationale is that we should not be encouraging scientists to think about science through awards but through having fun and the desire to do their best science. That should be enough. The result is exactly that—having a bit of fun and making some important points all at once. I hope you enjoy it as much as I did.

1. http://www.ploscollections.org/article/info:doi/10.1371/journal.pcbi.1003858

OPEN ACCESS

**Citation:** Roberts RJ (2015) Ten Simple Rules to Win a Nobel Prize. PLoS Comput Biol 11(4): e1004084. doi:10.1371/journal.pcbi.1004084

**Editor:** Philip E. Bourne, National Institutes of Health, United States of America

**Published:** April 2, 2015



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

## Introduction

It is remarkable how many students, young faculty, and even senior faculty hanker after a Nobel Prize. Somehow, they think that it is possible to structure their scientific careers so that the culmination will bring this much sought-after honor. Some even think that as a Nobel laureate myself, I may have the key to success—some secrets that I can share and so greatly improve their odds of success. Unfortunately, I must begin by disappointing everyone. There is only one path that should be followed. It is summed up in Rule 1, but some of the other Rules may prove helpful—or if not helpful, then at least amusing.

## 1. Never Start Your Career by Aiming for a Nobel Prize

Don't even hope for it or think about it. Just focus on doing the very best science that you can. Ask good questions, use innovative methods to answer them, and look for those unexpected results that may reveal some unexpected aspect of nature. If you are successful in your research career, then you will make lots of discoveries and have a very happy life. If you are lucky, you will make a big discovery that may even bag you a prize or two. But only if you are extraordinarily lucky will you stand any chance of winning a Nobel Prize. They are very elusive.

## 2. Hope That Your Experiments Fail Occasionally

There are usually two main reasons why experiments fail. Very often, it is because you screwed up in the design by not thinking hard enough about it ahead of time. Perhaps more often, it is because you were not careful enough in mixing the reagents (I always ask students if they spat in the tube or, more recently, were texting when they were labeling their tubes). Sometimes, you are not careful enough in performing the analytics (did you put the thermometer in upside down, as I once witnessed from a medical student whose name now appears on my list of doctors who I won't allow to treat me even if I'm dying?). These problems are the easiest to deal with by always taking great care in designing and executing experiments. If they still fail, then do them over again! But the more interesting reason that experiments fail is because nature is trying to tell you that the axioms on which you based the experiment are wrong. This means the dogma in the field is wrong (often the case with dogma). If you are lucky, as I was, then the dogma will be seriously wrong, and you can design more experiments to find out why. If you are really lucky, then you will stumble onto something big enough to be prizeworthy.

## 3. Collaborate with Other Scientists, but Never with More Than Two Other People

Collaboration embodies much of what is good about science and makes it fun. By bringing different sets of expertise to bear on a problem, it is often the key to making discoveries. However, if you think you are getting close to a big discovery, always keep in the back of your mind that there can only be three winners on the ticket for a Nobel Prize. Pick your collaborators carefully! But seriously, don't do as some have done and try to make a competitor of someone who would otherwise be an extremely valuable collaborator.

## 4. To Increase Your Odds of Winning, Be Sure to Pick Your Family Carefully

Seven children of Nobel Prize winners have gone on to win the Prize themselves, and four married couples have jointly won the Prize. Marie Curie and her husband, Pierre, won in Physics in 1903, while their daughter Irene with her husband, Frederic Joliot, won the Chemistry Prize in 1935. Carl and Gerty Cori won the Medicine Prize in 1947, and Alva Myrdal and Alfonso Robles won the Peace Prize in 1942. Lawrence Bragg shared the Physics Prize in 1935 with his father, William. Roger Kornberg (Chemistry, 2006) and his father, Arthur, (Medicine, 1959) both won. Aage Bohr (1975) and his father, Niels, (1922) both won the Physics Prize. Other father-son laureates are the Swedes Hans von Euler-Chelpin (Chemistry, 1929) and Ulf von Euler (Medicine, 1970) and Manne Siegbahn (1924) and Kai Siegbahn (1981), both in Physics. Briton Joseph John Thomson (1906) and his son George (1937) both won the Physics Prize. The only siblings to bask in Nobel glory were Jan and Nikolaas Tinbergen (Medicine, 1973) of the Netherlands. Jan won the first Prize awarded in Economics in 1969.

With a total of 586 Nobel Prize recipients in science during the 113 years since it was first awarded, these are impressive numbers, given a world population numbering at least 10,000,000,000 over the same period of time.

This rule is vividly illustrated last year (2014) by another married couple sharing the Nobel Prize in Physiology or Medicine.

## 5. Work in the Laboratory of a Previous Nobel Prize Winner

Many Prize recipients have benefitted greatly from the inspiration that this approach can bring. Sometimes just working at an institution with a previous Prize winner can be helpful. One prime example is the Medical Research Council (MRC) Laboratory in Cambridge, United Kingdom, where no less than nine staff members have won Nobel Prizes in either Chemistry or Physiology and Medicine, including my own personal hero Fred Sanger, who won the Chemistry Prize twice (1958, 1980), once for inventing protein sequencing and once for pioneering DNA sequencing. In between, he also invented RNA sequencing, but perhaps three Prizes was more than the Nobel Committee could stomach.

## 6. Even Better Than Rule 5, Try to Work in the Laboratory of a Future Nobel Prize Winner

This can be very beneficial, especially if you can be a part of the Prize-winning discovery. That has proven to be a very good strategy, but it is not always easy to spot the right mentor, one who will bring you that sort of success and then share the glory with you. The corollary of this strategy is not to work in the laboratory of someone who has already won but whom you think will win again with you on the ticket. This has yet to prove successful based on the previous double recipients named in Rule 5! It is much better to make sure that any big discoveries come from you after you leave the lab and are out on your own.

## 7. Always Design and Execute Your Best Experiments at a Time When Your Luck Is Running High

A casual survey of Nobel Prize winners will soon confirm that most credit luck as being the biggest component in their discovery. This is partly because many discoveries arise when what we think we know turns out to be wrong and we base our further research on incorrect assumptions. However, only rarely are we lucky enough to have to make such dramatic changes in our assumptions that a really major breakthrough becomes possible—the sort that may one day be considered appropriate for a Nobel Prize.

## 8. Never Plan Your Life around Winning a Nobel Prize

This has proven disastrous for many people. I know several scientists who became convinced that they were going to win and had all sorts of plans for less-than-modest speeches acknowledging the award of the Prize, preparing comments for journalists and planning subsequent trips to exotic places to talk about their discovery. It is far better not to know you have been nominated so that it comes as a real surprise when you get the early morning call from Stockholm. In fact, why not just forget about the Nobel Prize altogether and focus on doing the very best science you can? If you decide to ignore this rule, under no circumstances should you bug current Nobel laureates to nominate you. This has been an all-too-common strategy employed by many who feel they should be laureates, some even going so far as to send their last year's publications along every year with a reminder of what they consider their "big" discovery. This will almost guarantee that the laureate won't nominate you and is likely to lead to them advising their friends similarly. Can you imagine how that conversation would go after a few late-night drinks in the bar?

## 9. Always Be Nice to Swedish Scientists

Several laureates had their prize severely delayed by picking a fight with the wrong person, someone who was either already a Nobel Committee member or became one subsequent to the fight. Some individuals may even have lost out altogether, although one would need to search the archives (only available 50 years after the award) to find them. This is usually an easy rule to follow as in my experience the Swedes are very nice people, good scientists, easy to collaborate with, and extremely amiable drinking partners.

It is never too early to get started on this. Then, should your name magically appear on the candidates' list and you have to wait for it to reach the top, you may still be around to cash in. Peyton Rous had to wait from 1911 until 1966 for the Medicine Prize, just four years before his death.

## 10. Study Biology

There are many reasons for this. First, biology is fascinating, never boring, and directly affects our everyday lives, yet we still know relatively little about it. Thus, the odds of making a big discovery are greatly increased compared to other disciplines. Second, biology is all around us, is vastly complicated, and encompasses disciplines such as medicine, agriculture, conservation, and computer science, as well as many others, thus lending itself to the kind of interdisciplinary approaches that make science such fun and can easily lead into new territory. Third, unlike physics and chemistry, biology is ever changing, thanks to evolution. What seems to be the rule today may have changed by the time you are doing your experiments. Finally, there are two Prize categories in which biological discoveries are currently being awarded. One is Physiology or Medicine, and the other is Chemistry, in which about half the Prizes go to biologists. Already you have increased your odds by 50%.

## Conclusions

In summary, Rule 1 is the best advice I can offer. There is no substitute for pursuing the very best science that you can. Even Marie Curie, John Bardeen, and Fred Sanger needed this to win their second Prize. In contrast, Linus Pauling, one of the cleverest chemists of his generation, only received his second Prize (in Peace) by working in a totally different field. Nevertheless, the odds of winning a second Prize, if you already have one, do seem rather better than average!

## Acknowledgments

I would like to thank the following people who read drafts of this piece and offered helpful comments, as well as much advice: Tilde Carlow, Marty Chalfie, Dana Macelis, Shelly Glashow, David Lipman, Rahmah Mohamed, Chris Noren, Karen Otto, Mila Pollock, Janos Posfai, Ira Schildkraut, Barton Slatko, and, last but not least, my wife, Jean, whose comments cannot be repeated here.

Editorial

# Ten Simple Rules for Approaching a New Job

**Philip E. Bourne***

Office of the Director, National Institutes of Health, Bethesda, Maryland, United States of America

At some point in your professional career, you will be faced with a job interview. This may range from visiting a graduate school where you already have a placement should you want it, to interviewing for a very high-profile position in industry, government, or academia where there is significant competition for that job. Thinking both as a job applicant and a job interviewer about how I have approached job situations over the years before, during, and after the interview and how those situations have turned out, I can offer the following ten simple rules as you prepare. Where appropriate, I conclude a rule with an illustrative scenario for a junior- and/or senior-level position since while the general principles are universal, how they are applied depends somewhat on the seniority of the position.

## Rule 1: Really Want the Job

It is tempting to apply for a job even if you are not sure you want it. As an interviewer, I can say that reading a very generic job application sends a message that the person does not really want the job. This can waste a significant amount of your time as the applicant and the time of those conducting the job search. Chances are you will not get the position because that lack of want will be apparent during one or more interviews—assuming you get as far as an interview. You will lack the passion that the employers are looking for. Everyone, including you, will be disappointed. Be honest with yourself from the outset. Imagine yourself in the job two years in. Is it exactly where you want to be in your career—and life—in two years? Asking yourself whether you really want the job is particularly important if you have been approached to apply for the position. While this is gratifying, remember you are not the only one likely to be asked, and the askers will likely themselves benefit from your application. That benefit for them could be financial in the case of a headhunter approaching you, or more subtle, through improving the asker's reputation if you get the job. Obviously there is more to consider than just the job. A change of job is frequently a life-changing event as well, for example, through relocation, financial change, stress on the family, etc. Making plus and minus columns and discussing the potential job application with all those that it will touch is something that works for me. Then, imagine your life two years into the position and ask the appropriate questions of yourself. Imagine the case of your first tenure-track position, although similar questions apply universally: Am I being productive enough to get tenure? Do I like my work environment and my work colleagues? Am I happy living in this place? What are my future career prospects here?

## Rule 2: Wishful Thinking Is Not Enough—Be Qualified

It is tempting to apply for a position that you are not truly qualified for because you really want it (you have obeyed Rule 1), but deep down you know you are not qualified for it. Beyond the time wasting in applying for something which you have no hope of getting, there is the mental anguish associated with applying for a job. Time is spent wondering, "Will I get it, will I get it?" when that time could be used more productively. Before applying for a job, it is always a good idea to talk to mentors who will give you a candid opinion of your chances before you expend any effort. It may also be helpful to review the qualifications of those in similar positions to determine whether an application makes sense. Having said this, "being qualified" can be a qualitative term. Yes, there will likely be minimal degree requirements, but other aspects of the prerequisite requirements may not be so clear. Years of experience could substitute for a higher degree, relevant experience in a different field might count for something, and so on. Notwithstanding, deep down you will likely know whether you have a chance at a position—be honest with yourself. Again, imagine your first tenure-track application. Do I really have enough publications, grants or promise of grants, teaching experience, and proven service to get this job? Having said all this, it is possible you have a talent or experience that, while not identified in the job posting, really appeals to your potential employer. This is an unusual situation. Be realistic, but at the same time be ambitious—a balance that you will need to judge for yourself.

## Rule 3: Understand and Work the Process

Getting a new job is a process. There is the written application—including cover letter, CV, and possibly a vision or research interest statement of some kind—which you should have someone proofread. Submitting these materials will likely lead to a prescreening, and telephone and in-person interviews may follow. As a reviewer of many such applications, I have to say two things impress me. First, how well the skill set of the applicant maps to the position, and second, how much time the applicant has spent in tailoring the application for the particular position—including their CV. In my opinion, it counts for a lot if the applicant understands the work of the people they will be collaborating with and lays out specifically what they hope to accomplish in working with them. More obvious is the need for the applicant to conform to the process itself—if the application asks for a specific set of skills, outline those skills; more on that in the

**Citation:** Bourne PE (2014) Ten Simple Rules for Approaching a New Job. PLoS Comput Biol 10(6): e1003660. doi:10.1371/journal.pcbi.1003660

**Editor:** Fran Lewitter, Whitehead Institute, United States of America

**Published** June 26, 2014



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

* Email: philip.bourne@nih.gov

Philip E. Bourne is Founding Editor-in-Chief of *PLOS Computational Biology*.

subsequent rules. You will likely be asked for references as the process progresses. Choose these well. They will likely not be the people who will say the nicest things about you, but people whose opinions are most respected and can provide a value judgment against others in their network. Lastly, the selection process involves a very significant human factor. If, on paper, two applicants appear similar, the one that appeals most to the decision makers will invariably get the job. Think how you can best appeal to the decision makers. Know who those decision makers are (see Rule 8), and as far as possible, what they will be looking for in you as the applicant.

## Rule 4: Be Prepared—Have Something in Writing and Practice the Interview

This works for me and I think would work for most job applicants. Beyond the required documents, I like to map out in writing my thoughts about what I would contribute to the position that is not brought out in the formal application materials. This could be in the form of written answers to imaginary questions that are likely to arise during the interview. By thinking answers through and writing things down ahead of time, you will be less likely to give vague, trite, or at worst, wrong answers to important questions. Questions to address cover the details of the job itself and also questions that arise around many jobs relating to diversity, conflict of interest, ethics, etc. Even better if you can practice the interview with a colleague, or better still with an experienced interviewer. This gets you thinking on the spot and provides instant feedback on how you did. I would even consider videoing the mock interview for later review and diagnosis.

## Rule 5: Do Not Oversell Yourself

This applies both to the written application and any interviews but is more likely to be an issue in an interview situation when you are nervous and eager to impress. Quite simply, do not waffle, fib, or lie (obviously true of the written application too). If you do not know the answer to a question, or feel you do not have a particularly good answer to a question, then say so. While admitting to not knowing something, it is also a good time to indicate you are eager to learn and grow in the new position. Also, if you can't answer a question, request an answer from the interviewer; that will frequently lead to further discussion, which will likely readily indicate you know more than was first conveyed and, again, that you are collegial and willing to listen to the opinions of others.

## Rule 6: Do Not Undersell Yourself

If you come out of the interview thinking, "Damn, I forgot to mention so and so," then you were likely underprepared and undersold yourself, unless you happen to be well known to the interviewers. Rule 4 is helpful in this regard, since with proper written preparation you will be more likely to give a complete picture of your capabilities. So for example, be prepared to articulate exactly your contributions to your most important and most recent papers. Notwithstanding, in this preparation do not try and learn everything you will need for the job. Getting the job will not depend on what you crammed for the interview, but what experience and knowledge you have acquired over the proceeding years. Make sure that knowledge and experience comes across.

## Rule 7: Understand Your Potential New Workplace

This is important not just by way of helping you decide whether you want to go and work there—there is nothing worse than working in a toxic environment—but also in getting a job offer in the first place. It is all part of doing your homework for the position. This is more than just a web search. Use your network of colleagues to get a sense of the workplace. However, if those colleagues are in the institution to which you have applied, be careful not to put them in a compromising position. Having said that, if the opportunity arises, it is valuable to visit your potential new workplace and talk to people outside of the formal interview process. Let us use two specific examples to give this rule some perspective. First, you have a job interview as a new Assistant Professor in a university in a geographic region new to you. Visit the institution and wander around a day early if you can. Understand the institution—what is the student population, how is it distributed, what are the institutional strengths and weaknesses, etc. Understand the department and/or school you would become part of—what is the faculty to student ratio, what is the breadth of the syllabus taught, what is the research strength, what is the organizational structure, etc. Understand what you will be expected to contribute—suggested courses to teach, collaborative research to undertake, etc. Second, you have a job interview as a software engineer in a for-profit company. Be familiar with the products and services of the company, understand the competition, have some ideas of what you can contribute towards improving products or providing new products. Understand the management structure and how you would fit in.

## Rule 8: Understand Your New Colleagues

As an interviewer, I am impressed if the candidate knows something about what I do and how it relates to their application—what can I say, beyond that I am human. I have also seen this overdone, leaving me with the awful impression the job candidate had been stalking me. Like all that is presented here, there is a balance between overdoing and underdoing it; at least be familiar with the interviewer's latest papers. As an interviewer, if the applicant can see how they would fit in with a couple of specific examples, I will be pleased. Again using our Assistant Professor scenario, that would mean what you would like to teach that would complement courses already offered and a couple of specific research ideas that involve specific collaborations with members of the department and/or school you would be entering. As an interviewee, I ask myself, "Can I see myself working with these folks? How do their interactions and body language bode for my own future in this environment?"

## Rule 9: Be Both Assertive and Humble

This is another example of the need to achieve an imaginary balance that is hard to learn except by experience. It's a component of that nebulous part of your personality known as "people skills." As you advance in your career, this becomes less of an issue, as you by then have a reputation that is known to at least some of those interviewing you, which got you to the interview stage in the first place. Earlier in your career, you are more likely to be unknown and have got the interview on the strength of your written application and CV. In this case, people skills are important. At the very least, you need to leave the interviewer with the impression, "Yes I would like to hire this person." To me, that implies that the candidate is both gently assertive and, at the same time, humble. I can't begin to describe how to achieve this.

## Rule 10: Follow up

If there are outstanding issues from the application process, particularly the interview, it is wise to follow up with the chair of the interview committee or the individual interviewer. As an interviewer, this sends two messages to me. First, the candidate would seem to really want the job, and second, I have got additional information that will help in an informed decision. I do not like receiving gratuitous follow-up but rather meaningful input into the decision-making process that I did not have thus far. Others disagree, and believe any kind of follow-up thank you is appropriate.

These rules are just simple guides I have found useful. What should be clear is that this is one person's view, and I invite you to add your own comments on what has worked and not worked for you during the job interview process, either as a candidate or interviewer. Consider it a challenge to crowdsource the perfect job application.


## Acknowledgments

Thanks to Scott Markel for useful input and for hiring several of my students over the years! Thanks also to Anthony Fletcher for a business management perspective.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Internship in a Pharmaceutical Company

**Bin Chen***

Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, Stanford, California, United States of America

Students have to choose between industry and academia at the end of their formal education. One great opportunity to assess their ability in industry is to intern in a company during their academic training [1]. Many pharmaceutical companies often offer a 10- to 12-week summer internship program for undergraduate and graduate students. During this program, interns have an opportunity to work with dedicated mentors from the company on real-world industrial problems related to drug discovery and development as well as to participate in other training activities organized by the internship program office. Depending on the particular project and student performance, interns should be able to accomplish one or more short scientific projects. Some interns may be able to present the results of their work in the form of scientific papers or conference talks or even land a job through the training. This short-term working experience in pharmaceutical companies will help students better understand the pharmaceutical industry, learn the process of drug discovery and development, and build a strong network with experts and fellows in the pharmaceutical field, which can positively contribute to future career development. In addition, it will help students to identify if they really enjoy working in industry and help them in choosing a future career after school.

Thus, it is of extreme importance to make full use of this rare opportunity to explore the pharmaceutical field, to nurture yourself, and to prepare for a career after school. Many students often manage to have one or even more internship trainings during their school years. For those who are planning to do an internship, I would like to tell you about my five internship experiences in three big pharmaceutical companies (namely, Novartis, Pfizer, and Merck) during graduate school at Indiana University, Bloomington. The rules and advice also come from my former intern fellows, mentors, and colleagues from pharmaceutical companies. Besides computational biology and chemistry, many of these rules may be applied to other industrial fields as well.

## Rule 1: Start Preparing for an Internship Early

You need to start looking for internship postings and consulting seniors as early as possible. Submissions of internship applications usually start in December, and an offer is typically issued in March of the next year (the dates vary widely among companies). If you are enrolled in a Master of Science (MS) program and planning to intern in the first summer, you should start preparing right after orientation. You can search postings from job search engines (e.g., indeed.com and Monster.com), professional communities (for example, the International Society for Computational Biology (ISCB) for jobs for computational biology, CCL.NET for jobs for computational chemistry, and LinkedIn groups), or company websites. Some websites provide an email-alert service that sends new job postings automatically. Another great source in which to look for internships is a scientific conference organized by the American Medical Informatics Association (AMIA), Intelligent Systems for Molecular Biology (ISMB), the American Chemical Society (ACS), or Bio-IT World.

Early preparation not only gives you more opportunities to find a suitable position, but also lends you a much longer time to prepare your basic skills. If you start searching before December, you may not see job postings for the next year, but you would see those from the last year, which could guide you in skill preparation. Often the postings list basic requirements and preferred skillsets. Those do not change significantly from year to year. For example, cheminformatics-related programs may expect you to have some experience with particular software and tools (e.g., Pipeline Pilot and Spotfire), while bioinformatics-related programs may expect you to know a bit about public databases (e.g., Gene Expression Omnibus [GEO]) and sequence processing tools and pipelines. These skills can be extended through the courses you are/will be taking or through online programs. You do not need to be an expert, but broader experience is always a plus. Don't be afraid if you don't have extensive experience; these skillsets can be gained quickly, assuming you are not afraid of hard work and willing to learn new things.

## Rule 2: Leverage All Your Sources to Be Selected for an Interview

Interviewers often get a pile of resumes for one position, but only a few of them will be selected for an interview. Except for those resumes coming from prestigious schools or labs, most other resumes, particularly from fresh graduate students, do not have much to distinguish them. In that case, how do you make yours stand out? Other than making your profile perfectly match the job description, you can try networking, which is sometimes very helpful in getting an interview. If the interviewers happen to know you or your references, your chances will definitely increase.

As a new student with no connection to industry, what should you do? The best person to consult is your advisor, who knows you best and has many more connections. Many companies would send the job posting to your advisor and also

**Citation:** Chen B (2014) Ten Simple Rules for Internship in a Pharmaceutical Company. PLoS Comput Biol 10(5): e1003600. doi:10.1371/journal.pcbi.1003600

**Published** May 8, 2014



**Funding:** The author received no specific funding for writing this article.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: binchen1@stanford.edu

contact your advisor before they made a decision to hire you. Therefore, do make a good impression on your advisor and tell him/her ahead of time that you would like to apply for an internship. The seniors in your group or school are another valuable resource. Ask them (especially those who have interned before) if they know of any openings anywhere or know anybody who could forward your resume to your dream company. In addition, many conferences (e.g., ISMB) have student sessions that allow students and industrial fellows to mingle.

Another possible way to get an internship would be to contact company scientists directly if you know their work and it is in an area of your interest or expertise. They may be able to work through the internship office if they see a fit. They may even have the opportunity to influence the job posting.

## Rule 3: Survey the Interviewer(s) before and during Interview

An interview can be conducted any time after you submit your application. You may have two interviews for a program and often may have one week or more for preparing an interview. The manager or mentor usually conducts the first interview by phone, followed by another phone interview from human resources (HR) if the manager or mentor approves. Therefore, impressing the manager or mentor is key. In addition to being familiar with all the job requirements and each point in your curriculum vitae (CV), knowing your interviewer's background in advance can be very helpful. You may be able to get a glimpse from LinkedIn or from reading the interviewer's publications.

The phone interview is a great chance to learn more about the position and show your interest, curiosity, and excitement about the internship opportunity. Don't be afraid to discuss the projects you will be working on, the sources you may use, or any questions that will demonstrate your interest. If you are able to raise a couple of interesting questions on their recent work, they will definitely be very impressed.

## Rule 4: Specify Your Target at the Beginning of the Program

Congratulations if you get an offer, but keep in mind that three months is a really short time, especially if you want to accomplish an impressive project, so be sure to prepare in advance before you start. You should take one project first and focus on it. Your mentor may already have a detailed proposal or may only have a general idea. Whatever the situation is, at the beginning you have to sit down with your mentor to discuss your expectations of each other and your objectives and then come out with a practical way to achieve them. If you are not able to come out with a concrete plan immediately, literature review is a helpful way to start. Sometimes, you may want to propose an alternative plan if you can foresee some potential problems. It is recommended that you start with a project that may be publishable, as a publication definitely will benefit both you and your mentor.

## Rule 5: Keep to the Timeline

A 12-week program can be divided into three phases: 2 weeks, 6 weeks, and 4 weeks. The first 2 weeks would be taken up by the orientation, group introduction, training, and project planning. In the 9th or 10th week, you would start wrapping up your work, and in the following week, a poster session would be kicked off, so you basically only have about 6 weeks to devote to your work. Thus, you have to make a clear timeline at the beginning and note every day what you will do and what you have done so that your work is kept on track. It is suggested that you have a weekly meeting with your mentor and coworkers to discuss your progress and address any problems right away. Some companies may have a formal midterm evaluation, which can give you valuable feedback on your progress.

## Rule 6: Don't Hesitate to Ask Questions

Industry greatly values teamwork. Other than your mentor, there may be other colleagues involved in your project, so be ready to communicate with them. Since you basically only have 6 weeks or so to devote to your project, don't waste your time working on something that has been done already. Particularly, at the beginning you need many sources to jumpstart your project. Don't be afraid that you will look stupid by asking basic questions. Your goal is to understand the field and project better, work with your teammates to move forward more productively, and enjoy the progress of learning in the meantime. Your mentors should already have reserved some time to answer your questions, but if their schedule is very tight, you may either email your questions or ask for more meetings. They are always able to squeeze some time in for you.

## Rule 7: Have the Confidence to "Sell" Your Project

In big companies, collaboration between diverse groups happens quite often. Many companies even hold regular social events so that scientists from different backgrounds can mingle. Take advantage of these opportunities to talk about your work with others. As many of them don't know your work, being confident will help convince your audience to listen to your work and appreciate it. Some work—for example, developing a tool for analyzing microarray data—may not seem particularly novel to computational biologists, but for many bench biologists, this tool could save time significantly. Timely feedback may not only inspire your work, but may also help establish potential collaboration. Finally, a poster session offered at the end of the internship can be a good place to "sell" your project. Be sure to write down the names of the colleagues who are extremely interested in your work in case your mentor would like to follow-up.

## Rule 8: Expand Your Horizons beyond Your Project

Journal club, seminars, team meetings, intern training sessions, and many other activities organized by the education office provide opportunities to learn about the company and the process of drug discovery and development. Many companies organize a session in the middle of the summer to introduce the company and hold a social event to help interns communicate with company leaders and management teams. The group or department also has regular meetings that allow you to learn about the work of your colleagues. In addition, an internal e-source is extremely valuable in nurturing yourself. Many companies have their own internal wiki or SharePoint sites, from which you can learn about group projects without bothering your colleagues. You can even participate in drug discovery training for free through their e-learning sources. These are not required during an internship but definitely are benefits you should not miss.

## Rule 9: Be Social, Open-Minded, and Curious

Other than communicating with your mentor and coworkers, you should also be active in engaging with other colleagues, either at the lunch table or during casual talk. It is encouraged to make an appointment with other colleagues individually. Just simply say hi to them and inquire whether they have some time to have a chat. Despite the busy schedules of your colleagues, they often would be very happy to find some time for you in order to learn more about you, discuss their projects, and share their working experience. If you are living with other intern fellows, don't miss that social time either. Otherwise, try to find someone, talk with him or her, and share your experience. Be sure to acknowledge all the colleagues who help you either in your poster or in your lab presentation.

## Rule 10: Finishing the Program Does Not Mean the Ending

You will find that time is flying fast, and soon you have to wrap your work up. Due to the short-term working period, it may happen that you could not finish the project as planned, but you need to document your work in an accessible manner so that your colleagues can continue the work without much effort. Be sure to store the data, analysis, documents, and code as specified several days before leaving so that your colleagues have enough time to look at your work and ask you questions face-to-face. Some groups have a source code version control tool; be sure to test it a few times before you leave. If you think the work is publishable, be sure to talk with your mentor in advance and discuss the work that needs to be done before you leave. Once you leave the company, it's hard to run the program remotely or access the data.

Finishing the internship does not mean the end. In addition to working on a manuscript, you may need to work with your colleagues. By this point, your professional network should be established. Other than being connected on LinkedIn, you should contact your mentor or colleagues occasionally and report your research progress and career plan. If you are interested in working in industry, definitely let them know that you are looking for opportunities. Some interns may land a job after the internship or benefit from the internship during their job hunting. More importantly, you should extend this network to the new students in your group or your school and let the network continue to grow.

## About the Author

**Bin Chen** is a postdoctoral scholar at Stanford University. During his graduate school at Indiana University, Bloomington, he completed five internships (including one as a contractor) at Novartis, Pfizer, and Merck.

## Acknowledgments

The author is grateful for the support from Atul Butte during his postdoctoral training, the support from David Wild during his graduate school, and the helpful comments from Aleksandra Markovets, Abhik Seal, Hui Huang, Lun Yang, and Robert Sheridan.

## References

1. Searls DB (2009) Ten simple rules for choosing between industry and academia. PLoS Comput Biol 5: e1000388. doi: 10.1371/journal.pcbi.1000388.

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules to Protect Your Intellectual Property

**Mark Jolly[1], Anthony C. Fletcher[2], Philip E. Bourne[3,4]***

**1** Brookes Batchelor LLP, London, United Kingdom, **2** Salix Management Consultants Ltd, London, United Kingdom, **3** Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America, **4** Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, United States of America

The concepts that underpin the protection of ideas and inventions are not new; such laws have been around for several hundred years and are discussed under the broad heading of intellectual property (IP). IP is easily misunderstood, but at the same time most scientists encounter it at some point in their career, as it is a necessary feature in the commercialization of research.

The term intellectual property includes such concepts and rights as copyright, trademarks, industrial design rights, and patents. It is important to remember that IP is a tool to help your endeavours, and not a goal in itself. Having IP for its own sake is pointless. IP can be crucial in commercializing research and running a successful science-based business, but having a patent and having a successful patented product are two very different things.

Above all, IP can only work for you if you understand what it is, why you want it, and what you are going to do with it. These ten simple rules are intended to provide an overview of these issues; however, we must start with a warning. Laws relating to IP change all the time, they are complex, sometimes rather obscure, and are very different from country to country. For example, research surrounding methods of treatment by surgery and therapy and diagnostic methods are patentable in the United States, but specifically excluded from patentability in Europe [1]. However, these boundaries seem to be shifting in both the US and Europe. In short, we are dealing with a complex and changing subject and restrict ourselves here to the guiding principles.

## Rule 1: Get Professional Help

Although the process of obtaining IP looks deceptively simple, like many things the devil is in the detail. Let's consider patents as an example. The practicalities of patent application are straightforward; you simply file documents with the relevant body indicating that a patent is sought, and provide the identity of the person applying and a description of the "invention" for which a patent is sought. The patent office will then write back to you with an application number.

However, there is no guarantee that a patent application will become a granted patent. Indeed, at the application stage they do not even check that your description describes an invention at all. Even if you draft a description in as much detail as you would for an academic research paper and file it yourself, the prospect that it will be granted and enforceable is very low. There is skill and technique, even a language, that patent attorneys and patent agents have that allows them to describe and define inventions in the way a patent office requires. As an example, in everyday parlance, the terms "comprise" and "consist" could be considered to mean the same, whereas they have very distinct meanings in a patent application.

The dangers are possibly even greater with trademarks and registered designs (also known as "design patents")—these are generally granted with very little examination and patent offices are often even less inclined to suggest using a patent/trademark attorney for such "simpler" rights; however, the lack of examination means the validity of such a right is uncertain and they become open to challenge.

The costs of redrafting a self-filed application are invariably higher than the costs for drafting an application from scratch, and if there has been any disclosure it will probably not be possible to re-draft. So, in summary, if you want your IP to be valuable, you should seek professional advice at an early stage.

## Rule 2: Know Your (Intellectual Property) Rights

IP rights come in various guises, and each is a defensive right to pursue legal action in the event that a third party infringes. In very basic terms:

- Patents protect inventions—broadly, things that are new and not obvious—and the way they work. Sometimes this is expressed as "everything under the sun made by man"; however, there are numerous local exceptions from patentability—we touched on the complexities of methods of treatment above—but there are similar issues in relation to genes, computer programs, and business methods, for example.
- Registered designs protect the appearance of products (not the function, which is protected by patents).
- Trademarks protect brands (e.g., trade names and logos).
- Copyright protects the expression of ideas—i.e., the words you choose to use to describe your idea—not an idea itself.

Most businesses do not need the trinity of patents, trademarks, and designs; in fact, trademarks are probably the only IP most companies have or need, however for a few companies the full house is required: for example, consider the Apple® iPad®: two registered trademarks, a registered design for its shape, and of course patents for the way it interacts with the user. Not to mention copyright covering the code and the packaging. A huge battle in the courts around the world is currently taking place over these rights that may well effect changes in the law. The *Wall Street Journal* calls the recent Apple/Samsung case "the patent trial of the century" [2].

**Citation:** Jolly M, Fletcher AC, Bourne PE (2012) Ten Simple Rules to Protect Your Intellectual Property. PLoS Comput Biol 8(11): e1002766. doi:10.1371/journal.pcbi.1002766

**Published** November 8, 2012



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: pbourne@ucsd.edu

Philip E. Bourne is Founding Editor-in-Chief of *PLOS Computational Biology*.

## Rule 3: Think about Why You Want IP (i.e., What You Will Actually Do with It)

Any money spent on IP is capital that cannot be spent on production, marketing, etc., so think carefully about why you are investing in protecting your IP. There are many good reasons: to stop people from copying you; to add value to your company if you want to sell it; to sell or license to a third party; to hold it in your armoury if you suspect you are going to be sued and want to countersue (for example, Google has spent a substantial amount of money buying patents recently [3]); even to reduce your tax bill (in certain countries profits attributed to patents can be taxed at a lower rate [3,4]).

However, in general, IP is a right to prevent other people from doing something; owning IP does not necessarily give you the right to do anything yourself.

One school of thought says that IP is only valuable if you are willing to enforce or defend it, and the cost of such an action can be prohibitive. Indeed, the business model of "patent trolls" is to purchase patents, sometimes from those who cannot afford to enforce them, not to use the invention, but just to enforce against infringing companies. On the other hand, the term "defensive IP" has been used to describe IP obtained, not to stop other people from competing, but to stop a competitor from patenting something that you may wish to use in the future. Thus a patent application may be filed, and published but allowed to lapse, with no intention of ever enforcing it, simply because the step of publication will mean that should a competitor apply to patent the same or a similar invention, the patent office will locate your application and it will anticipate the competitor's application.

Note also that while this article is titled "Ten Simple Rules to Protect *Your* IP", it is important not to be too introspective and to consider other people's IP. For example, successful strategies can be built around taking exclusive licenses—licenses that exclude even the IP owner from using the IP. One tactic to improve your competitive position can be to take an exclusive license under a patent, then either expand your range to include the patented product, or continue only to sell your own product, but use the exclusive license to prevent manufacture of the other by anybody else.

## Rule 4: If You Don't Protect the IP, Your Innovation Is Less Likely to Happen

Maybe you are not an entrepreneur yourself, but have an idea that you would like to see it exploited—it could, after all, make the world a better place. You can publish it—then anyone who wishes can use it freely. But the big question here is, will they? Many inventors think that by publishing their ideas freely they are more likely to have them exploited; however, the converse is often true (for example, in health care, where lack of patent protection is often cited as a major reason for not following up an idea (T. Roberts, former president of the Chartered Institute of Patent Attorneys [UK]).

The reason is economic: most innovations require investment, and investors look for a return on their money. However, ideas that are released without any IP protection will often immediately attract competitors who can perhaps undercut the inventor (for example, with economies of scale). This decreases the likelihood of investment in the development of an invention (which is often more crucial than the invention itself) and increases the need for investment in marketing, etc. to obtain a competitive edge.

So what we have to consider here is that—even if you don't want to profit personally from the innovation—it may still pay to protect it so that it will see the light of day through other investors. Remember, IP can be licensed and what happens to the resulting income is up to the IP's owner. And this is a point where it gets complex for scientists and others who invent as part of their employment. We will cover this in more detail in Rule 10.

## Rule 5: What's in a Name?

You have a great idea but it's not patentable, or you have applied for patent protection but are worried that it may not cover everything, and of course the protection will expire after 20 years [5]. This is where trademarks come in to fill the gap in your protection. Unlike patents and designs, a trademark or brand can be protected with a registration at any time (unless someone else has got there before you)—you do not need to have kept your name a secret, and once registered the right will only expire if you stop using it or fail to renew it (generally every 10 years). So, you can protect your invention with a patent and sell it under your brand, which is also protected. Once the patent protection expires, customers are used to buying your product with reference to your brand, and will hopefully continue to do so even though competitors may start offering rival products. Just make sure your brand is something memorable and unique to you.

Viagra is just one example of a trademark so closely associated with the product (sildenafil) that a good proportion of the market should remain in the hands of the trademark owner well after the patent has expired (in this instance, if priced competitively). You do need to be careful here in selecting the name you are protecting: descriptive brands are easy to market but hard to protect because descriptive terms do not fulfil the requirement of "distinct character". And you can be too successful: many people now use the trademark Hoover to mean a generic vacuum cleaner, Thermos for a vacuum flask to keep food hot, or Tannoy for a public address system. It can be very expensive in terms of lawyers fees to police such trademarks and keep protecting these names and prevent them becoming simply part of the language and hence devalued.

## Rule 6: Be Realistic about What You Can, and Cannot, Protect

IP rights are, generally speaking, national rights provided by individual governments to regulate activity in that particular country. In some cases there are bilateral and multilateral agreements (for example, most of the world has signed up to the Berne Agreement, which accords the same level of copyright protection to foreign nationals of other Berne states that is provided to nationals of the state concerned [6]).

However, for most rights, it is a national issue. In an ideal world, each incremental improvement would be patented in each national jurisdiction (there are approximately 200 countries in the world), along with the name you trade under, and every brand would be the subject of a trademark, as would any color associated with your company and any sound you use, your products and their packaging would be the subject of registered designs, and your patent attorneys would be very wealthy!

In the real world it is essential to be realistic. A patenting regime covering more than the US, Europe, and a handful of other countries is a rare sight outside the realms of very large companies (such as big pharma), and even many big companies restrict themselves to key markets.

## Rule 7: It's Big Business and Controversial

The world of IP is a big one. It's controversial, as it has a huge impact on international relations and trade. It's also controversial for political reasons, as many people feel that aggressive protection stifles the utility of products that have the potential to do good in the emerging world (again, for example, big pharma). The World Intellectual Property Organization (WIPO) is the United Nations agency dedicated to this area [7], and it's worth considering its overarching aims, which include reducing the knowledge gap between developed and developing countries, and ensuring that the IP system continues to effectively serve its fundamental purpose of encouraging creativity and innovation in all countries.

Of course, many question the value to society of IP, or at least the expansion of IP, in promoting creativity and innovation. The Public Library of Science describes itself as a driving force of the open-access movement, and accordingly, unlike many copyrighted works, this article may be copied without seeking permission, provided that the original authors and source are cited.

It can be hard, for example, to defend the extension of copyright from 50 years after an author's death to 70 years on the grounds that the extra 20 years of protection is in any way likely to encourage creativity. Whatever your thoughts on IP, it is worth bearing in mind that others may disagree.

As a scientist and innovator you may be driven by many ideals: to make the world a better place, perhaps, or to buy yourself a yacht—we are all different. But like it or not, if you want to commercialize your ideas you cannot avoid the issue of IP, and we go back to Rule 1 here—get professional advice. Even if your aim is totally philanthropic you may still need to invest to protect your innovation, perversely because this is what will give it the biggest chance of actually succeeding. Simply make sure you tell your patent attorney what your ultimate aims are.

## Rule 8: Keep Your Idea Secret until You Have Filed a Patent Application

Little upsets a patent attorney more than hearing "I have a great idea—it's selling really well" or "I've shown it to a few companies and they seem very interested".

There is an old maxim that says a secret shared is not a secret anymore. While a secret shared under a non-disclosure agreement (NDA)—documents most people have heard about but probably never read—ought to stay secret, discussing an invention under the umbrella of confidentiality is no substitute for being able to freely discuss or publish an idea that is protected by a patent application.

Obviously, once your idea is published by a journal it is too late to file a patent application—your invention has been made available to the public. However, earlier in the publication cycle the situation is different. If you send a paper to a journal for submission, it will (excluding open review) be treated as a confidential disclosure to the publisher and the reviewers. Notwithstanding, the best advice is still to file a patent application before submitting a paper, either to avoid a potential "abusive disclosure" or hold up the publication of the paper.

In summary, novelty is key to patentability and your own disclosures count against you, so remember to file a patent application *before* telling anybody who is not bound by confidence.

## Rule 9: Trade Secrets

Regarding patents, the economic reasoning behind the system is an exchange between you and the public. The government allows you a monopoly, and your side of the bargain is to disclose fully your invention so that once your 20 years of protection is up, it can be freely exploited for the good of society. A patent can provide you with a 20-year government approved monopoly. However, some ideas cannot be patented and indeed, some innovators don't want to patent their ideas. All is not lost here, however, as we fall back on an older idea and one much beloved of thriller writers: the trade secret.

If you really can keep a secret, your monopoly on an idea or product may never end. But once the genie's out of the bottle, like a champagne cork, you won't get it back in and you are unlikely to extract sufficient damages from whoever breaches confidentiality. Thus, if you have an idea that cannot be reverse engineered, you do not have to enter into the patent bargain. Trade secrets are free—just prevent the secret being disclosed. But bear in mind that that this can be very difficult indeed, but not impossible. Famous successful examples include the recipe for Coca-Cola and the formulation of the alcoholic beverage Chartreuse, which is only known by two monks.

## Rule 10: Make Sure the IP Is Owned in a Way That Allows Development

Notice that we don't suggest "make sure *you* own the IP of your invention". If you discover something whilst working as an employee (e.g., of a company or an academic establishment), there will certainly be something in your contract about this. Generally, the employer will have first call on the invention, but may have clauses that will return rights to the individual if it is not exploited within a certain time—in some countries this is enshrined in law [8].

Ownership of IP is a minefield, and can be particularly difficult in an academic setting where numerous complicating features are involved. Universities, as employers, are likely to have a right to their employees' inventions; funding bodies may make their own claim; inventorship is not like authorship—the people whose names are on an academic paper are unlikely all to be inventors; and in cross-border collaborations, national laws on ownership may well be in competition with each other. One complicating factor that is often encountered is joint ownership: if you can, avoid joint ownership; instead, set up a company to own the IP and license it to partners if necessary (otherwise you face differing national rules on what joint owners can do with and without each other's permission).

If it is necessary to share IP, work out at the beginning who owns what, what rights each party has and importantly who will have the right to future inventions. In fact this is a common theme in several of our Ten Simple Rules: as soon as money rears it ugly head, strife follows, so it's as well to plan for dispute resolution right from the beginning.

In summary, first, you can never act too early, but it's very easy to act too late. Like many topics that involve the law, IP is a mind-numbingly complex topic and more so, perhaps, as it's not national, but international, so get the very best professional advice you can. If you are working as an employee, speak to your company at the earliest stage; they have a vested interest in helping get it right. Second, because significant sums of money are involved, plan for future discord. Finally, persevere: your invention can make the world a better place.


## Acknowledgments

Grateful thanks to Dr. Tim Roberts and Clare Turnbull (both of Brookes Batchellor LLP) for their invaluable help in preparing this article. Thanks to David Searls and Scott Markel for their suggestions during review.

## References


1. The European Patent Convention. Article 53. Exceptions to patentability. Available: http://www.epo.org/law-practice/legal-texts/html/epc/2010/e/ar53.html. Accessed July 2012.
2. Jones A, Vascellaro JE (2012) Apple v. Samsung: the patent trial of the century. July 24, 2012. Available: http://online.wsj.com/article/SB10000872396390443295404577543221814648592.html. Accessed July 2012.
3. BBC News (2012) Google adds IBM patents as it looks to future. BBC News. January 4, 2012. Available: http://www.bbc.co.uk/news/technology-16409081. Accessed July 2012.
4. HM Revenue & Customs (n.d.) Corporation tax reform: patent box. Available: http://www.hmrc.gov.uk/budget2012/tiin-0726.pdf. Accessed July 2012.
5. The 20 year period is virtually harmonized the world over thanks to the Trips agreement: Available: http://www.wto.org/english/tratop_e/trips_e/t_agm0_e.htm. Accessed July 2012.
6. World Intellectual Property Organization. Summary of the Berne Convention for the Protection of Literary and Artistic Works (1886). Available: http://www.wipo.int/treaties/en/ip/berne/summary_berne.html.Accessed July 2012.
7. World Intellectual Property Organization. Available: http://www.wipo.int. Accessed July 2012.
8. Deck M, Matthes J (2005) Employee inventions in Germany. Intellectual Asset Management. June/July 2005. Available: http://www.iam-magazine.com/issues/article.ashx?g=e22454bd-0d83-46a8-9e25-96386aaefafd. Accessed July 2012.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules to Commercialize Scientific Research

**Anthony C. Fletcher[1], Philip E. Bourne[2,3]***

**1** Salix Management Consultants Ltd, London, England, **2** Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America, **3** Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, United States of America

Commercializing scientific research or a breakthrough idea is really no different, in principle, from commercializing anything, except perhaps that it's more difficult in practice because of the steps required to turn basic research into something practical and because you are looking for a market for a product, rather than designing a product to fit an established, or obvious market.

Commercialization is different to starting and running a company, a broader endeavour and the subject of a previous Ten Simple Rules article [1]. Even so, commercialization can be a broad endeavour. For example, at one extreme, you could hand over your monoclonal antibody to Sigma to supply it on your behalf to other researchers who might find it useful while the company pays you a small royalty; on the other, you could be involved in developing Herceptin (anti-HER2 monoclonal antibody) from its origins as a mouse-specific antibody through to its use as an effective anti-breast cancer drug, in a process that took more than decade. Here we assume the former—others are carrying out that commercialization, which has its pluses and minuses—less work for you, but typically less control of the commercialization process.

Commercialization is a much studied subject, both by academics [2] and the business community [3]. All larger academic institutions generally have offices to promote and help scientists get research to market. Consequently, in this Ten Simple Rules article we won't deal with the details, but instead will concentrate on some of the key issues to consider when working with, or before and after working with, a specialized office.

## Rule 1: What Drives Science Does Not Drive Business

Scientists evaluate research by considering whether it makes an original contribution to our understanding of the world. Businesses have a different rationale, which, by and large, is to make money. This engenders a huge culture gap. In the 18th century, as the Chinese started to make porcelain for European markets, it was noted that they simply didn't get the idea of perspective. Pagodas appeared the size of flower vases. The artists understood symbolism; Europeans sought realism. And so it is with commercialization: scientists are not primed for business (some would even say this goes against academic freedom) and businesses are not, for the most part, so good at science unless they have specialized research divisions—Bell Labs comes to mind here, although these days an exception rather than a rule. When these worlds collide there is a need for intermediaries and translators to ensure a common understanding and successful path from research to commercialism. Scientists need to get business people who are "on the same wavelength" on their team and who can explain and guide them. Conversely, businesses have to be able to determine what research universities have to offer and how it could be of benefit. Interfaces are varied, ranging from university development offices to business outreach units to organizations like CONNECT (http://www.connect.org) that specialize on being the interface. These are valuable resources and should be utilized by both scientists and potential business partners.

## Rule 2: There Is No Single Path to Commercialization

Commercialization of scientific breakthroughs is something that has become more formalized in recent years thanks, in the United States at least, to the Bayh-Dole Act (legislation dealing with intellectual property arising from federal government–funded research) [4], with academia taking an active role in facilitating the translation of its intellectual capital into business. There are many routes for this: licensing, royalties, incubation, and in-house development. Industry itself has also moved physically closer to large universities (e.g. science parks) to share in the human capital. Beneath all this activity there are complex issues regarding how much potential value lies locked up in these intellectual assets and how they can best be developed without straying too far from the progenitors' ideals, and at the same time generating value. There are many ways to go from the laboratory bench to the store: commercialization is just like any business process– part art, part science; part inspiration, part perspiration. Most routes are essentially mechanistic, some work and some don't—there is no secret way to do things. So if anyone tells you at the start it's a sure fire winner (or not), don't believe them—there is a lot of hard work that has to be done to see if an idea can make it. And never believe advice that says "this is the best way" based on a single example—for every research-driven idea that makes it big, hundreds wither slowly away. These failures are hardly ever the subject of detailed case studies, and so we have no idea why they failed and what lessons we could learn.

## Rule 3: You Must Know Your Rights and Those of Colleagues

This might sound obvious, but it is important to know who owns and who has the right to develop your research output. As academics, by default, most institutions (or less often, funders) own your research. The institution may choose to protect your ideas with copyrights, licenses, or patents, a wise idea if they are to have commercial value (see Rule 4). That protection is not on your behalf as the inventor, but on behalf of the institution(s) where the work

**Citation:** Fletcher AC, Bourne PE (2012) Ten Simple Rules to Commercialize Scientific Research. PLoS Comput Biol 8(9): e1002712. doi:10.1371/journal.pcbi.1002712

**Published** September 27, 2012



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: pbourne@ucsd.edu

Philip E. Bourne is Founding Editor-in-Chief of *PLOS Computational Biology*.

was done. You need to understand what this protection means in terms of process, cost, and time involved. Research is collaborative, often with multiple institutions involved, and this can greatly complicate the rights and ownership of intellectual property. Such issues should have been thoroughly reviewed and agreed with all the relevant scientists before the research is disclosed. Good scientific collaborations can be ruined by misunderstood commercialization strategies.

## Rule 4: Consider the Implications of Going from Public to Private

Academic research has many benefits, for example, collaboration, data and knowledge sharing, and freedom to publish. When moving this research into the private sector, different rules apply. There is a need to protect the intellectual property. In some cases, protecting that investment has implications for follow-on developments and impacts academic freedom. For example, consider a situation where a company licensing a technology from an academic institution also has the rights to follow-on developments. Those rights could impact the academic scientist's ability to freely publish those new developments.

## Rule 5: Decide How Much of Yourself You Want to Give

At one extreme, you can give over your research completely and have little or nothing to do with subsequent commercialization; at the other extreme you could be heavily involved in the company commercializing your research or indeed found a company to develop the research. The level of engagement with the commercialization is going to define the time commitment and possibly financial reward coming from the commercialization. This needs to be thought about carefully at the outset and should be mapped to your longer-term career goals. Some academics want to, and do, make a successful transition to business—perhaps as happy heads of research and development (R&D), free from the administrative hassle, but a key part of the business—and some of course stay in academia. Markets have no sentiment and don't care what you do: they just care what you can contribute.

## Rule 6: Separate the R and the D and Be Realistic

There is a big difference between basic research and the development of such research to the point of commercialization. Generally, development is done by the entity commercializing the product and could be considered the mid-point between academic and commercial cultures. Development can be hugely expensive and time-consuming and presents a huge financial risk to the investor, especially as it is a front-loaded cost. The investor has to look at such topics as mass production (scaling up from lab levels), distribution, logistics, pricing, practicality, marketing, safety, the law, etc. Often times, one or more of these proves insoluble and the breakthrough has to languish, possibly for decades, until a solution appears. Personal genomics is an example where extensive commercialization of a number of ideas has had to wait until next generation sequencing makes the products feasible. Scientists also need to be realistic in valuing the idea—they typically have no concept of the development costs and often feel the basic research represents the bulk of the value, which is almost never the case.

## Rule 7: The Market May Not Exist at the Outset

The old fashioned method of working out what your factory can make (being "production led" in the jargon) and then seeing if there is a market is a largely discredited approach in modern business. In the case of basic scientific research, of course, this is exactly the situation—scientists invariably investigate things out of intellectual curiosity without any view to commercialization. The original research will not be aimed at solving any commercial, market-related problems, outside of obvious areas such as pharmaceuticals and engineering, and so the breakthrough is inevitably made in isolation of market requirements.

There are various anecdotes that illustrate the apparent lack of market. "Who needs music on the move?" was one comment about the Sony Walkman. "No one wants a tablet computer with no keyboard", and so on. Examples like these are often used to "prove" that a good idea will make it anyhow, but it's simply not true in the majority of cases. It conveniently sidesteps the point that if no ready market exists, it has to be developed. That takes money, advertising, skill, and time. All of which add to the development costs.

## Rule 8: Consider the "Want" versus the "Need"

There is a venerable marketing axiom that products should always address a need, not a want. People often express "wants", but they buy "needs". Consumers want a Ferrari but they buy a Toyota. It is so easy for an academic scientist to believe there is a need for a product resulting from their research when in fact it is a want (or to put it another way, it's a "nice to have" not a "must have"). Thus, commercialization of a breakthrough needs to address what people or other businesses will actually pay for—and this is a complex issue. Generally, a fair amount of time and money needs to be spent on market research to understand this—if people will not pay, then no matter how good the idea, it will never be successfully commercialized. Other market dynamics can also intervene: for example, a common issue is that of technologies that are never implemented because their payback time is greater than a market will bear. Market-related short-termism has killed many a promising idea.

## Rule 9: Make It Comprehensible

The people who are going to fund the development of your research and subsequently take it to market will be business people, not scientists, irrespective of whether the ultimate product is aimed at technical buyers. At the earliest stage you need to boil down the research into an "elevator pitch"—a few sentences the layperson can comprehend and one that sets out a clear reason to purchase. A mini reactor that fits in a suitcase and will power a domestic car for a year without recharging fits that model. A new aerogel does not—who knows what this does and what benefits it might confer? A common problem is that the relationship of the research to the final practical product may not be clear. One approach to solve this is by association: "Our breakthrough is a distinct improvement on…" Focus on the biggest profit opportunities in your early pitches. Business people prefer to see a clear track to a clear market opportunity rather than have to work it out for themselves.

## Rule 10: Customers Are the Ultimate Peer Review

As scientists, peer review of our research publications evaluates novelty, a correct and accurate scientific process, reproducibility, and value to the community. The example of Henri Poincaré is useful here to illustrate the value of peer review: the first version of his work on "The three-body problem" contained a serious error that was picked up during peer review.

## About the Authors

**Anthony C. Fletcher** has a PhD in organic chemistry, and had a 20-year career as a management consultant with Deloitte and with Cap Gemini Ernst and Young and is now a freelance consultant.

**Philip E. Bourne** has a PhD in physical chemistry, is a professor of pharmacology and Associate Vice Chancellor for Innovation and Industrial Alliances at the University of California San Diego, the co-founder of SciVee.tv, and founding editor-in-chief of this journal. The authors have known each other since childhood and have been involved in one company together.

Alterations and changes then led to extremely important work on modern chaos theory. Consider also Frege's pioneering book on predicate logic at the turn of the century. Bertrand Russell read it in draft form and sent him a letter pointing out it was prey to Russell's Paradox ("the set of all sets not containing themselves…"), and Frege was able to add a note in proof acknowledging this and discussing ways out. In business, the analogy is the importance of testing out ideas and products before a full launch and then to listen carefully to what the ultimate consumers say. This market research is key; if the market is lukewarm, it doesn't matter how great the research, a product won't happen. You need to be prepared for the eventuality that while the market research does not indicate a product can arise as you envisioned, a different product might be possible. Is that what you want?

As we said at the outset, looking for a problem to fit your solution is always going to be tough going. And it's probably even tougher to find someone who will back you with money, time, and resources that will be needed to turn your scientific research into something that will benefit society. But don't give up. Post-it Notes were once a scientific curiosity, Teflon just flakes in a solution and penicillin contamination in a petri dish.

Do remember, however, that as the originating scientist, knowledge and recognition may be the only reward you get—others who take it to market (and take the financial and commercial risk) might get the majority of the money. But as an academic scientist, hopefully that's not why you entered science in the first place. There is increasing emphasis worldwide for making better practical use of fundamental scientific research from academia. Be part of the change.

## Acknowledgments

Thanks to John Barber and Tony Ferrari for their insightful advice and to Theo Bloom, Ruth Nussinov, Scott Markel, and David Searls for their comments and input on the manuscript.

## References

1. Fletcher AC, Bourne PE (2012) Ten simple rules for starting a company. PLoS Comput Biol 8: e1002439. doi:10.1371/journal.pcbi.1002439
2. MIT Sloan School of Management (2012) Faculty & research: Fiona Murray. Available: http://goo.gl/tb1HB. Accessed 20 August 2012.
3. Isis Innovation Ltd. Available: http://www.isis-innovation.com. Accessed 20 August 2012.
4. Wikipedia contributors (2012) Bayh–Dole Act. Wikipedia, The Free Encyclopedia. Available: http://en.wikipedia.org/wiki/Bayh%E2%80%93Dole_Act. Accessed 20 August 2012.

Editorial

# Ten Simple Rules for Starting a Company

**Anthony C. Fletcher[1], Philip E. Bourne[2,3]***

1 Salix Management Consultants Ltd, London, United Kingdom, 2 Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America, 3 Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, United States of America

Many faculty, staff, and students at academic institutions think about starting companies at some point in their careers. As academic funding models change, and how academia views entrepreneurial activity changes, starting companies is likely to happen more frequently. Hence, it is worth considering *Ten Simple Rules* to contemplate when starting a company while in academia. There is a wealth of general information out there to help you, but that information is not aimed specifically towards computational biologists. What follows is a hybrid that is intended as a general quick review for anyone, intermingled with some specific advice for computational biologists.

By way of experience, we should say we have been involved in starting several companies: both in the biomedical sciences, dealing with biological software, computational biology services, and currently SciVee Inc. (http://www.scivee.tv) distributing scientific rich media, and outside it, ranging from the distribution of independent films, to a socially oriented dining club aimed at supporting local businesses, to a business supplying art quality photographic prints. None have been a great financial success, but all have been immense fun, an opportunity to meet interesting people, and an opportunity to think quite differently than when doing scientific research. Read that as a personal endorsement to go for it, even if starting a company is not yet a well-formed idea, and even though, as you will see below, the rules themselves might be cautionary and off-putting.

## Rule 1: A Great Product or Service Alone Is Not Enough to Make a Successful Business

This is a general rule of business. Very rarely is a product alone unique and desirable enough to make a successful business, even though the majority of academic founders tend to think so at the beginning. To misquote Ralph Waldo Emerson, "Make a better mousetrap and the world will beat a path to your door" is perhaps one of the most misleading axioms ever circulated. It is useful to continuously remind oneself that at least nine out of ten companies fail: indeed, most VCs (venture capitalists) will tell you that most business ideas they see should be canned before they ever reach the business plan stage, let alone before serious money is spent on development. Success comes from a huge amount of hard work, extensive market research, a realistic business model, and, above all, the application of good all-around business skills to all aspects of the business. For example, if you are great at sales but poor at money management, your business will certainly fail. Notwithstanding, in one sense the process of starting a business is not that different from academic research (replace business plan by grant and think about current grant funding levels), but it is the way the process is executed that is alien to most academics. Hence Rule 2.

## Rule 2: Business Is Part Art, Part Science

Success as an academic comes from having an eye for detail and contemplating a problem from every conceivable angle and only moving forward when you believe you have fully analyzed the problem. Businesses are typically different—the ecosystems involved are very complex, so they cannot be tested or researched to anything like the same degree of certainty. On occasion you must act quickly and based only on gut instinct. It is not by chance that economics is referred to as the "dismal science". There is a degree of luck behind every success.

## Rule 3: Define to Yourself and Others Why You Are Starting a Company and Be Prepared for Those Reasons to Change

There are many reasons academics start companies. At one end of the spectrum is the desire to make money; at the other end is the desire to do something for humanity. For most of us the reason is usually somewhere in between. That sounds fine, but in practice the time will come when one has to make hard choices and sacrifice one for the other. At the outset, define for yourself and others who are working on establishing the company where each of you stands on this issue, what is motivating you, and what you are willing to sacrifice. Remember also that a company is a legal entity in its own right and exists only to serve its stakeholders and no one else. As a rule of thumb, it often takes three CEOs to take a company to floatation—one to start it, one to develop it into a thriving business, and one to float or sell it—all different skills. Boards and executives will come and go and companies will change over time as they have to adapt to a market and not the other way round. Consequently, once you have breathed life into a company it can continue, with or without you, and its shape and purpose may change radically as markets demand.

## Rule 4: Decide What of Yourself You Are Willing to Put into the Company

There are many forms of investment to start companies—ranging from your own credit card or bank loan, angels (private investors), institutions, or VCs. It is a rare business that can be started on its own internal cash flow or is unique in its offerings to the marketplace. A common metric used by investors as part of evaluating a company is to ask the founders what money and time they are willing to invest in the company. The answer is telling to them and to you. Most

**Citation:** Fletcher AC, Bourne PE (2012) Ten Simple Rules for Starting a Company. PLoS Comput Biol 8(3): e1002439. doi:10.1371/journal.pcbi.1002439

**Published** March 29, 2012



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: pbourne@ucsd.edu

Philip E. Bourne is Editor-in-Chief of *PLoS Computational Biology*.

academics will not give up their day jobs to establish a company, which is a red flag for investors unless you can immediately compensate for this issue with the required resources, such as an experienced CEO or other team members. Hence Rule 5.

## Rule 5: Get Professional Business Help Early

Starting a company will require many skills an academic does not typically have. The general outline of a business plan describes these amongst others:

- Product/manufacturing/service delivery
- Manufacturing process
- Sales and marketing
- Business management and administration
- IT
- HR/personnel
- Financial management
- Legal, patents, contracts, etc.

The needed depth of such skills will also vary depending on the maturity of the business.

You can learn some of these skills as you go and there is fun in that, but for some tasks you will undoubtedly need deep expertise that can only be had from business professionals: for example, don't write your own contracts—get a lawyer to review them and don't expect to go from your laboratory and become a competent and successful salesperson. Above all, get a decent bookkeeper or accountant on board, as nothing is more depressing than spending your evenings doing the accounts and invoicing! So, recognize that you will need help and that help has to be paid for in some way—either through equity or more commonly, actual money—your shares in a start-up don't pay the weekly bill at the supermarket! There are consultants and professional firms that specialize in helping start-ups, and of course most are upright and professional. But not everyone is, so get some advice from business professionals you trust before settling on a course of action that may be irreversible.

## Rule 6: Understand the Legal, Ethical, and Regulatory Environment Thoroughly

Companies operate in bureaucratic and litigious environments unfamiliar to most academics. The key here is to get the right advice early on—you won't be able to proceed without lawyers, accountant, insurers, and many other professionals. But, remember their interests and yours are not necessarily aligned even though you are their client. So the rule here is to do your own research first and understand what is a "must have" and what is simply an "optional extra". You can then operate in a pragmatic way and keep your expenditure of time and resources to a reasonable minimum. Finally, remember that investors will expect you to have covered the bases properly and be operating on advice from recognized experts, which is usually expensive.

## Rule 7: Establish a Relationship with Your Academic Institution at the Outset

You need to understand the relationship you as an institutional employee and the company you are founding have with the academic institution in which you work. Typically, if the product comes from your laboratory, the institution in which you work owns the intellectual property (IP) and the company needs to license it from the institution. The work that produced the IP, and perhaps the associated product(s), was likely funded by a third party, and their relationship to the institution and to the company needs to be established. Typically, a government funding agency will confer ownership on the institution in which the work was done. Where things get tricky is when the IP and/or associated product(s), or follow-on IP and product(s), are developed in the company, what is your institution's rightful share? This is where the lawyers or other negotiators come in to establish IP rights.

## Rule 8: Realistically Define the Value of Your Business

It is easy to believe that the initial value of your product, and hence company, is much greater than it is. We all read of those rare cases where a start-up is sold to Google or Microsoft for a small fortune, but they are the exceptions. For example, in computational biology, a company based solely on software has little to no value. Your heart and soul may have gone into developing the software, but the hard fact is that others can likely recreate it, and it is difficult to protect; therefore, to an investor it has little value. What does have value is your experience in using that software to achieve the desired outcome, which is a service business with typically a limited but targeted audience. On top of that, the software must be of commercial quality that requires quality assurance, professional technical writers, a support/help desk, etc. Professional marketers call this the unique selling proposition—it's what separates you from your competitors and why people will buy from you—sometimes it's the product, sometimes the service, and sometimes the people, or a combination of all things. Whatever it is, it must be clear and compelling, and then it's a straightforward matter to put some metrics around it, work out what it's worth, and see who is interested.

## Rule 9: Think about Conflict of Interest Every Day

This is a serious issue that could impact your standing in the university and your community of scientific peers. Your professional standing will likely always be more valuable to you than being a co-founder of a company, so be careful. Many institutions have a conflict of interest office that can assist you in making the right decisions. The bottom line is that nothing you do in the laboratory should be driven by what can be perceived as being for the sole benefit of the company. Obviously, there is a huge gray area when it benefits the laboratory and likely the company as well. A useful exercise before undertaking something that might be perceived as a conflict is to conjure up how someone who didn't like you could spin what you are doing as a negative article in a reputable newspaper. How does that article read? Academic institutions competing for students and research funds do not like bad publicity.

## Rule 10: Decide Responsibilities and Equity Share before You Start

The nicest people and the best collegial relationships will be tested when forming a company together. The farther you go in establishing the company, the more vested each of those involved will become in its success. This is good on the one hand, but very bad if expectations for reward are not set at the outset. When incorporating the company, an initial set of shares is assigned; this is the time to define the initial equity share and should be based on past contributions to the enterprise and what each founder and employee will contribute to the company. Shares that vest over time (i.e., potential value that increases the more time and effort you put into the company) are a typical means of defining on-going contributions as well as monetary rewards. We would go so far as to say that in the life of a company there are always falling outs: the advantage of a limited company is that company rules typically state what should happen in such circumstances and the shareholders decide collectively. So never go ahead without planning for some acrimony and never ever go ahead on the basis of a handshake or as an unwritten partnership—lawyers are expensive in resolving conflicts.

As we said at the outset, the rules are a cautionary tale, but do not be deterred. Businesses are hard work but they are fun; you'll enjoy the experience. Nothing really beats the handshake (and signing) that closes a sale, or the unsolicited testimonial that tells you what you produced helped someone. You'll meet lots of people you would never have met before and probably go to places (nice and not so nice) that you would not go as an academic. You will take pride in having created something tangible that people value; they paid money for it after all. But remember never to compromise your academic principles—they are the most valuable asset you have.

## About the Authors

**Anthony C. Fletcher** has a PhD in organic chemistry, and had a 20-year career as a management consultant with Deloitte and with Cap Gemini Ernst and Young, and is now a freelance consultant.

**Philip E. Bourne** was trained as a physical chemist, is a professor of Pharmacology at the University of California San Diego, and the co-founder of SciVee.tv and editor-in-chief of this journal. The authors have known each other since birth, still go motorcycling together, and have been involved in two companies together.

Editorial

# Ten Simple Rules for Building and Maintaining a Scientific Reputation

**Philip E. Bourne[1]*, Virginia Barbour[2]**

**1** Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, California, United States of America, **2** Public Library of Science, Cambridge, United Kingdom

*While we cannot articulate exactly what defines the less quantitative side of a scientific reputation, we might be able to seed a discussion. We invite you to crowd source a better description and path to achieving such a reputation by using the comments feature associated with this article. Consider yourself challenged to contribute.*

At a recent Public Library of Science (PLoS) journal editors' meeting, we were having a discussion about the work of the Committee on Publication Ethics (COPE; http://www.publicationethics.org/), a forum for editors to discuss research and publication misconduct. Part of the discussion centered on the impact such cases have on the scientific reputation of those involved. We began musing: What on earth is a scientific reputation anyway? Not coming up with a satisfactory answer, we turned to a source of endless brainpower—students and other editors. Having posed the question to a group of graduate students, PLoS, and other editors, we got almost as many different answers as people asked, albeit with some common themes. They all mentioned the explicit elements of a reputation that relate to measurables such as number of publications, H factor, overall number of citations etc., but they also alluded to a variety of different, qualitative, factors that somehow add up to the overall sense of reputation that one scientist has for another.

What these students and editors identified en masse is one important side of a scientific reputation that is defined by data; but they also identified a much more nebulous side, that, while ill-defined, is a vital element to nurture during one's career. A side defined to include such terms as fair play, integrity, honesty, and caring. It is building and maintaining this kind of less tangible reputation that forms the basis for these Ten Simple Rules. You might be wondering, how can you define rules for developing and maintaining something you cannot well describe in the first place? We do not have a good answer, but we would say a reputation plays on that human characteristic of not appreciating the value of something until you do not have it any more.

A scientific reputation is not immediate, it is acquired over a lifetime and is akin to compound interest—the more you have the more you can acquire. It is also very easy to lose, and once gone, nearly impossible to recover. Why is this so? The scientific grapevine is extensive and constantly in use. Happenings go viral on social networks now, but science has had a professional and social network for centuries; a network of people who meet each other fairly regularly and, like everyone else, like to gossip. So whether it is a relatively new medium or a centuries-old medium, good and bad happenings travel quickly to a broad audience. Given this pervasiveness, here are some rules, some intuitive, for how to build and maintain a scientific reputation.

## Rule 1: Think Before You Act

Science is full of occasions whereupon you get upset—a perceived poor review of a paper, a criticism of your work during a seminar, etc. It is so easy to immediately respond in a dismissive or impolite way, particularly in e-mail or some other impersonal online medium. Don't. Think it through, sleep on it, and get back to the offending party (but not a broader audience as it is so easy to do nowadays with, for example, an e-mail cc) the next day with a professional and thoughtful response, whatever the circumstances. In other words, always take the high road whatever the temptation. It will pay off over time, particularly in an era when every word you commit to a digital form is instantly conveyed, permanently archived somewhere, and can be retrieved at any time.

## Rule 2: Do Not Ignore Criticism

Whether in your eyes, criticism is deserved or not, do not ignore it, but respond with the knowledge of Rule 1. Failure to respond to criticism is perceived either as an acknowledgement of that criticism or as a lack of respect for the critic. Neither is good.

## Rule 3: Do Not Ignore People

It is all too easy to respond to people in a way that is proportional to their perceived value to you. Students in particular can be subject to poor treatment. One day a number of those students will likely have some influence over your career. Think about that when responding (or not responding). As hard as it is, try to personally respond to mail and telephone calls from students and others, whether it is a question about your work or a request for a job. Even if for no other reason, you give that person a sense of worth just by responding. Ignoring people can take other serious forms, for example in leaving deserving people off as paper authors. Whether perceived or real, this can appear that you are trying to raise your contribution to the paper at the expense of others—definitely not good for your reputation.

**Citation:** Bourne PE, Barbour V (2011) Ten Simple Rules for Building and Maintaining a Scientific Reputation. PLoS Comput Biol 7(6): e1002108. doi:10.1371/journal.pcbi.1002108

**Published** June 30, 2011



**Funding:** No funding was received by Philip E. Bourne for this work. Virginia Barbour is paid a salary by the Public Library of Science, and she wrote this editorial during her salaried time.

**Competing Interests:** Virginia Barbour's individual competing interests are at http://www.plosmedicine.org/static/editorsInterests.action.

* E-mail: pbourne@ucsd.edu

Philip E. Bourne is Editor-in-Chief of *PLoS Computational Biology*. Virginia Barbour is Chief Editor of *PLoS Medicine* and Secretary of COPE.

## Rule 4: Diligently Check Everything You Publish and Take Publishing Seriously

Science does not progress in certainties—that is one of its joys but also what makes it such a hard profession. Though you cannot guarantee that everything you publish will, in 50 years' time, be shown to be correct, you can ensure that you did the work to the accepted standards of the time and that, whether you were the most junior or senior author, you diligently checked it (and checked it again…) before you submitted it for publication. As a first author you may well be the only one who appreciates the accuracy of the work being undertaken, but all authors have a responsibility for the paper. So, however small or big your contribution, always be upfront with your co-authors as to the quality and accuracy of the data you have generated. When you come to be a senior author, it is so easy to take a draft manuscript at face value and madly publish it and move on. Both actions can come back to haunt you and lead to a perception of sloppy work, or worse, deception. As first author, this mainly lets down your other authors and has a subtle impact on your growing reputation. As the senior author of an error-prone study, it can have a more direct and long-lasting impact on your reputation. In short, take publication seriously. Never accept or give undeserved authorship and in addition never leave anyone out who should be an author, however lowly. Authorship is not a gift—it must be earned and being a guest or gift author trivializes the importance of authorship. Never agree to be an author on a ghostwritten paper. At best these papers have undeclared conflicts of interest; at worst potential malpractice.

## Rule 5: Always Declare Conflicts of Interest

Everyone has conflicts of interest, whether they are financial, professional, or personal. It is impossible for anyone to judge for himself or herself how their own conflict will be perceived. Problems occur when conflicts are hidden or mismanaged. Thus, when embarking on a new scientific endeavor, ranging from such tasks as being a grant reviewer, or a member of a scientific advisory board, or a reviewer of a paper, carefully evaluate what others will perceive you will gain from the process. Imagine how your actions would be perceived if read on the front page of a daily newspaper. For example, we often agree to review a paper because we imagine we will learn from the experience. That is fine. Where it crosses the line is when it could be perceived by someone that you are competing with the person whose work you are reviewing and have more to gain than just general knowledge from reviewing the work. There is a gray area here of course, so better to turn down a review if not sure. Failure to properly handle conflicts will eventually impact your reputation.

## Rule 6: Do Your Share for the Community

There is often unspoken criticism of scientists who appear to take more than they give back. For example, those who rarely review papers, but are always the first to ask when the review of their paper will be complete; scientists who are avid users of public data, but are very slow to put their own data into the public domain; scientists who attend meetings, but refuse to get involved in organizing them; and so on. Eventually people notice and your reputation is negatively impacted.

## Rule 7: Do Not Commit to Tasks You Cannot Complete

It tends to be the same scientists over and over who fail to deliver in a timely way. Over an extended period, this becomes widely known and can be perceived negatively. It is human nature for high achievers to take on too much, but for the sake of your reputation learn how to say no.

## Rule 8: Do Not Write Poor Reviews of Grants and Papers

Who is a good reviewer or editor is more than just perception. Be polite, timely, constructive, and considerate and, ideally, sign your review. But also be honest—the most valued reviewers are those who are not afraid to provide honest feedback, even to the most established authors. Editors of journals rapidly develop a sense of who does a good job and who does not. Likewise for program officers and grant reviews. Such perceptions will impact your reputation in subtle ways. The short term gain may be fewer papers or grants sent to you to review, but in the longer term, being a trusted reviewer will reflect your perceived knowledge of the field. Although the impact of a review is small relative to writing a good paper in the field yourself, it all adds up towards your overall reputation.

## Rule 9: Do Not Write References for People Who Do Not Deserve It

It is difficult to turn down writing a reference for someone who asks for one, even if you are not inclined to be their advocate; yet, this is what you should do. The alternative is to write a reference that (a) does not put them in a good light, or (b) over exalts their virtues. The former will lead to resentment; the latter can impact your reputation, as once this person is hired and comes up short, the hirer may question aspects of your own abilities or motives.

## Rule 10: Never Plagiarize or Doctor Your Data

This goes without saying, yet it needs to be said because it happens, and it is happening more frequently. The electronic age has given us tools for handling data, images, and words that were unimaginable even 20 years ago, and students and postdocs are especially adept in using these tools. However, the fundamental principle of the integrity of data, images, and text remains the same as it was 100 years ago. If you fiddle with any of these elements beyond what is explicitly stated as acceptable (many journals have guidelines for images, for example), you will be guilty of data manipulation, image manipulation, or plagiarism, respectively. And what is more, you will likely be found out. The tools for finding all these unacceptable practices are now sophisticated and are being applied widely. Sometimes the changes were done in good faith, for example, the idea of changing the contrast on a digital image to highlight your point, but one always needs to think how such a change will be perceived and in fact whether it might, even worse, give the average reader a false sense of the quality of that data. Unfortunately, even if done in good faith, if any of these practices are found out, or even raised as a suspicion, the impact on one's career can be catastrophic.

In summary, there are a number of dos and don'ts for establishing a good reputation—whatever that might be. Do not hesitate in giving us your thoughts on what it means to be a reputable scientist.

Editorial

# Ten Simple Rules for Getting Ahead as a Computational Biologist in Academia

**Philip E. Bourne***

Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, California, United States of America

Getting a promotion or a new position are important parts of the scientific career process. Ironically, a committee whose membership has limited ability to truly judge your scholarly standing is often charged with making these decisions. Here are ten simple rules from my own experiences, in both getting promoted and serving on such committees, for how you might maximize your chances of getting ahead under such circumstances. The rules focus on what might be added to a CV, research statement, personal statement, or cover letter, depending on the format of the requested promotion materials. In part, the rules suggest that you educate the committee members, who have a range of expertise, on what they should find important in the promotion application provided by a computational biologist. Further, while some rules are generally applicable, the focus here is on promotion in an academic setting. Having said that, in such a setting teaching and community service are obviously important, but barely touched upon here. Rather, the focus is on how to maximize the appreciation of your research-related activities. As a final thought before we get started on the rules, this is not just about you, but an opportunity to educate a broad committee on what is important in our field. Use that opportunity well, for it will serve future generations of computational biologists.

## Rule 1: Emphasize Publication Impact, Not Journal Impact

Reviewers who do not know your work well, unless told otherwise, will often judge that work primarily by the journals in which it appears. If the majority of your papers are in *Nature* and *Science*, then let the system continue to fool the reviewer. For the rest of us, it is important to emphasize that the impact of the journal does not necessarily reflect the impact of your paper. Include any data that reflect the value of your work regardless of the journal. The number of times the paper has been cited and the download statistics for that paper are obvious metrics, but should be put in context. A few citations and downloads do not necessarily mean the paper is not valuable in a narrow field. Tell the committee why it has significant impact in that field. There are also other less likely sources of support that can help. Coverage by the Faculty of 1000, press releases, blogs, and any positive commentary on the paper by others are also valuable indicators of impact.

## Rule 2: Quantify and Convince

Reviewers may not be that familiar with the concept of article-level metrics and what they say about your science—where applicable, convince them in your application. Let me use an example. The very first article I wrote in this series was titled "Ten Simple Rules for Getting Published" [1]. It has been downloaded over 65,000 times, which is about 35 times per day since it was published 5 years ago. At the same time, according to Google Scholar it has been cited 30 times and according to ISI Web of Knowledge 11 times. The implication is that it has had some scholarly impact that is not reflected by the more traditional citation metric. In this case, the scholarly impact is mainly pedagogical in that it assists in professional development. This is easily overlooked by a promotion committee, but of some value in academic promotion. Metrics may not tell the whole story, for instance, in work that is relatively new. Use your application to inform the reviewers why you believe your work is significant.

## Rule 3: Make Methods and Software Count

Keep statistics on software and methods use. For example, keep statistics on the number and diversity of users of the software, publications that cite the software, and the impact of those citations. For software that is modular, include the diversity of applications to which those methods and/or software have been applied. Describe what it took to develop the methods and/or software and what impact that has on the community. Many reviewers will not appreciate what it takes to develop and maintain methods and/or software for the community. Do what you can to help the reviewer with details of your time and resources, and that of others, in maintaining the software for the good of the community. Educate the committee on what open source implies, assuming your software is open source. Indicate as best you can how your efforts in software and methods bring credit to the institution.

## Rule 4: Make Web Sites Count

This follows from Rule 3, but applies specifically to Web sites where Google Analytics, AWStats, and other tools can be used to quantify the impact your work has had and present those statistics to reviewers. Another irony is that papers about Web sites are rarely read, but they are highly cited if your resource is useful. Hence, they can be used to enhance your standing. Good professional conduct should dictate that you only write such papers when you have something substantively new to report regarding improvements to the Web site. Spreading citations over multiple papers just to enhance your H-factor while not adding anything substantively new speaks poorly of you and to the value system we use to evaluate scholars.

**Citation:** Bourne PE (2011) Ten Simple Rules for Getting Ahead as a Computational Biologist in Academia. PLoS Comput Biol 7(1): e1002001. doi:10.1371/journal.pcbi.1002001

**Published** January 6, 2011



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author is Co-founder of SciVee Inc., which works with Thomson Reuters on products. At this time, these products are not related to those outlined in this article.

* E-mail: pbourne@ucsd.edu

## Rule 5: Make Data Deposition, Curation, and Other Related Activities Count

Maintain records on your data-related activities, namely public accessibility, how much curation and other effort went into providing these data, and how much these data are used. Currently, there is no way to quantify the impact your public contributions of data have had on science; therefore, try to ensure that such contributions have an associated publication. Contact data resources to see if they can provide metrics for how frequently data you have contributed has been accessed and include that information in your list of accomplishments.

## Rule 6: Use Modern Tools to Emphasize/Quantify Your Academic Standing

Increasingly, tools are available to impart to reviewers your scholarly standing. For example, ResearcherID from Thomson Reuters [2] will provide graphs on the total number of citations per year, average number of citations per article, and so on. However, these are only for publications found in ISI databases, which can be limited for a multidisciplinary researcher. PubNet [3] will provide your collaborative network from PubMed where each node on the network is a researcher you have published with and the thickness of edges reflects the number of times you have published together. BioMedExperts [4] provides similar data. Again, this can be somewhat limiting for multidisciplinary researchers. Bolster these statistics by indicating the full range of your scholarly activities not covered by the tools. Adding papers manually to the tracking resource can often help as well.

## Rule 7: Make an Easily Digestible Quantified Summary of Your Accomplishments

Reviewers are often faced with many applications for promotion to review, and your accomplishments are easily lost in a long CV. This is particularly true if the reviewer is trying to sort out what you have accomplished in a specific time frame, as would often be the case when considering a promotion. One way to summarize accomplishments is as a bulleted list in a cover letter or some other allowable personal statement. Items on that list should include, where appropriate: published and accepted papers, pending and funded grants, including the amount coming to your institution, summarized accomplishments in software, data, and methods as per Rules 3, 4, and 5, students mentored and in what capacity, courses offered and their standing, other educational and outreach activities, company involvement, professional activities (e.g., editorial boards, scientific advisory boards), invited lectures, and awards. The idea is not to provide details here—your CV should do that—just numbers for easy and quick comprehension.

## Rule 8: Make the Reviewers' Job Easy

Often, one or more of the reviewers looking at your application are going to be responsible for writing a summary of why, or why not, your advancement was granted. Again, unless the reviewers are very familiar with your work they will appreciate a candid, quantitative and honest discussion of your accomplishments. But take heed of Rule 10. Where such a discussion should be included depends on the form of your application—usually as a cover letter or part of your personal statement is appropriate. Whatever the form, it should be brief and highlight, in a way that can be understood by a non-expert, what was done and why it is of high impact and, if available, how others have followed up on the accomplishments. These highlights should be peppered with citations and quantitative data that a reviewer can easily reference should they choose to do so. More often than not the reviewer will appreciate this summation and it will be reflected in the letter they write.

## Rule 9: Make the Job of Your References Easy

Often your application will include letters of support from external references, some chosen by you, others chosen by the reviewers. For the ones you choose, send those references the same summary you provide the reviewers (Rule 8). The reviewers will likely know your work well, which is why they were chosen. Notwithstanding, a good factual summary can help in their writing a reference letter, which is a significant undertaking when done well. They will thank you for it. You might even include information they would appreciate, that the committee would not—for example, specific details of research if you and the reviewer are in the same field.

## Rule 10: Do Not Oversell Yourself

This may be obvious, but have an impartial third party look over your application and have them give you a candid opinion; perhaps a senior member of your institution not on your committee. Don't oversell yourself with flowery adjectives. Show, don't tell; that means, enumerate facts. If you head a laboratory, even though it is your file under consideration, it is really the work of the collective you are highlighting—be clear and fair about that. Just state the facts—if you have done well, you will do well. It is as simple as that.

I have placed significant emphasis on what to include in a cover letter or personal statement that accompanies your CV, research statement, and perhaps other materials, such as teaching evaluations. I have not discussed preparing a good CV since such information is available on the Internet and elsewhere already. What has not been covered before, as far as I am aware, is how a computational biologist in academia might maximize their chances of being promoted by a committee that is not fully appreciative of the field.

As always, we welcome your comments. I would particularly like to hear additional/alternative advice from those like myself who have been through this process a number of times. In closing, I can only offer an example of such materials that I think helped me get promoted last time around (see Text S1).

## Supporting Information

**Text S1** Example support letter.
Found at: doi:10.1371/journal.pcbi.1002001.s001 (PDF)

## Acknowledgments

Thanks to my fellow Editors Barbara Bryant and Steven Brenner as well as previous Ten Simple Rules co-authors Jenny Gu, Quentin Vicens, and Iddo Friedberg for their important input.

## References

1. Bourne PE (2005) Ten simple rules for getting published. PLoS Comp Biol 1: e57. doi:10.1371/journal.pcbi.0010057.
2. Thomson Reuters (2010) ResearcherID home page. Available: http://www.researcherid.com/. Accessed 1 December 2010.
3. Douglas SM, Montelione GT, Gerstein M (2005) PubNet: a flexible system for visualizing literature derived networks. Genome Biol 6: R80.
4. Collexis (2010) BioMedExperts home page. Available: http://www.biomedexperts.com/. Accessed 1 December 2010.

PLoS COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Choosing between Industry and Academia

**David B. Searls***

One of the most significant decisions we face as scientists comes at the end of our formal education. Choosing between industry and academia is easy for some, incredibly fraught for others. The author has made two complete cycles between these career destinations, including on the one hand 16 years in academia, as grad student (twice, in biology and in computer science), post-doc, and faculty, and on the other hand 19 years in two different industries (computer and pharmaceutical). The following rules reflect that experience, and my own opinions.

## Rule 1: Assess Your Qualifications

If you are a freshly minted Ph.D., you know that you will need a good post-doc or two before you can be seriously considered for a junior faculty position. If you're impatient, you might be thinking of industry as a way to short-circuit that long haul. You should be aware that companies will strongly consider your post-doctoral experience (or lack thereof) in determining your starting position and salary. While you may not relish extending your indentured servitude in academia, any disadvantage, financial and otherwise, can quickly be made up in the early years of your career in industry. In other words, trying to get off the mark quickly is not necessarily a good reason to choose industry over academia.

On the other hand, you may have completed an undergraduate or Master's program with a view to going to industry all along, with never a thought of an academic career. You should still consider the point of the previous paragraph. While abbreviated "practical" bioinformatics training programs can be excellent, a Ph.D. is a significant advantage in all but the most IT-oriented positions in industry, at least at the outset. This is not to discourage anyone from embarking on a fast-track-to-industry program if their heart is in it, but be aware that the further you climb the educational ladder, the higher and faster you can start when you step across to the business ladder, and the better you will compete for a job in the first place. The days are long past when bioinformaticists were in such short supply that any qualification would do.

If you are an old hand and have already notched up a post-doc or two, take stock of your star power. This unspoken but universally understood metric encompasses such factors as whom you've trained with, where you've published (and how much), and what recent results of yours are on everyone's lips. If you are fortunate enough to have significant capital in this department, then the world may be your oyster, but you still need to consider where you will get the greatest leverage. While your stardom may be less taken for granted in industry, my feeling is that academia is a better near-term choice in such circumstances. Consider that it was in academia that you achieved the success you own thus far, so you obviously "get it." The simple fact is that academia is rather more of a star system (as in Hollywood) than is industry.

Finally, if you count among your qualifications a stint in industry already, as an intern or perhaps as part of a collaboration, you will not only be in a better position to compete for a permanent job, but you will be much better prepared to make the decision facing you. Stated another way, if you are seriously considering industry as a career path, you should probably have already taken advantage of the many opportunities out there to dip your toes in the water.

## Rule 2: Assess Your Needs

In taking stock of your needs, and perhaps those of your family, a decent living is generally at or near the top of the list. Salaries are still higher in industry, though the gap is not nearly so wide as it once was. If you need a quick infusion of cash, companies may offer signing bonuses, though again these were more common when bioinformatics was a rarer commodity.

Industry offers forms of compensation unavailable in academia, and you will need to consider how to value them relative to your present and future needs. Despite recent bad press, bonus systems are often part of the equation, and depending on your entry point they may constitute a significant percentage of total compensation. There is a tendency among academics to discount bonus programs in their comparison shopping, sometimes to zero, and this is a mistake. Bonuses are considered core aspects of compensation in most companies, and though they always have a performance-based multiplier, the base levels have historically been fairly dependable. That said, these are tough times in industry, and there are no guarantees. Your best strategy is to understand the reward system thoroughly, ask for historical data, and avoid comparing only base salaries unless you are extraordinarily risk-averse.

Share options are another matter. While in the past these were very attractive, and fruitful in practice, most industry types will tell you frankly that any options they've received in the past decade are deep underwater and a deep disappointment. Many consider pharma shares (and therefore options) to be a bargain at the moment, but that's between you and your financial adviser to assess. In any case, it is not a short-term consideration, since options typically take several years to vest.

If you are looking at biotech, however, share options and similar ownership schemes need to be a key consideration,

**Citation:** Searls DB (2009) Ten Simple Rules for Choosing between Industry and Academia. PLoS Comput Biol 5(6): e1000388. doi:10.1371/journal.pcbi.1000388

**Published** June 26, 2009



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: david.b.searls@gmail.com

David B. Searls is an Associate Editor of *PLoS Computational Biology.*

since these are a major rationale for assuming risk—more on that below.

Finally, you may have more specific needs to consider, such as a spouse also in need of a job. The two-body problem has always been tougher in academia than in industry, and probably always will be. If you are both academics, note that industry often has good contacts with local universities, and can facilitate interviews. Being a star certainly helps, so don't be afraid to negotiate. In fact, a general rule of thumb is that it never hurts to make your specific needs known, within reason. Academia will try to accommodate them as a community, while on the other hand business (particularly large, diversified companies) may have resources to address them that you wouldn't have expected. Nobody wants to hear a peremptory demand, but if a company wants you, be sure to let them know anything that might offer them a way to attract you.

## Rule 3: Assess Your Desires

There are needs, and then there are desires. Do you want riches? Fame? A life at the frontiers of knowledge? The hurly-burly of the business world? How do you really feel about teaching, publishing, managing, interacting, traveling, negotiating, collaborating, presenting, reporting, reviewing, fundraising, deal-making, and on and on? Though it may seem obvious, this is a good time to decide what really drives you.

First, the obvious. Do you want to teach? If lecturing is in your blood, your decision is made, although if a smattering will suffice you may have the option from within industry of an adjunct academic appointment. (By the same token, if you are not so enchanted with lecturing, grading, tutoring, etc., there are often options for research track professorships that minimize teaching duties.) Do you want to publish? While it will always be "publish or perish" in academia, it is certainly possible to grow your CV in industry, and it can even enhance your career, depending on the company. However, it might be largely on your own time, and you will likely encounter restrictions in proprietary matters, though in practice you can generally find ways to work within them. Ask about publication at the interview, both policies and attitudes, and watch out for any defensiveness.

An important question, surprisingly often overlooked, is how you want to actually spend your time, day by day and hour by hour. In academia, you will immediately be plunged into hands-on science, and your drivers will be to start out on your career by getting results, publishing, networking, and building your reputation with a view to impressing your tenure committee. A career in industry may put more of an early emphasis on your organizational aptitude, people skills, powers of persuasion, ability to strategize and execute to plan, etc.; in terms of growing your reputation, your audience will be the rather narrower community of your immediate management. A somewhat more cynical view would be that in business you will spend seemingly endless hours in meetings and writing plans and reports, while in academia you will spend all that time and more in grantsmanship—in this regard, you must pick your poison.

Finally there is the elephant-in-the-room question: Do you want to make money, or to help people? This is, of course, a false dichotomy, but many people consciously or unconsciously frame the decision in just this way, and you had best deal with it. Try thinking of it not so much in terms of the profit motives of the respective institutions, but in terms of the people with whom you would spend your career. You should have encountered a good sampling of scientists from industry during meetings, internships, collaborations, interviews, etc. (or in any case you should certainly try to do so before making judgments). If you are left in any doubt as to their ethics or sincere desire to relieve human suffering as efficiently as possible, or if you feel these are somehow trumped by the corporate milieu, then by all means choose academia—but only after applying analogous tests to the academics you already know well. In my experience, business doesn't have a monopoly on greed, nor are humanitarian impulses restricted to academia. That said, in the final analysis you must be comfortable with your role in the social order and not finesse the question.

## Rule 4: Assess Your Personality

Not surprisingly, some personality types are better-suited to one environment or the other. Raw ambition can be viewed as unseemly in either case, but there is more latitude for it in industry, and greater likelihood of being recognized and rewarded sooner if you are "on the go." In fact, one of the clearest differences between academia and industry are their respective time constants. Although the pace of academia may have quickened of late, it is still stately by comparison with industry, and much more scheduled (so many years to tenure, so many months to a funding decision, etc.). If you are impatient, industry offers relatively fast-paced decision-making and constant change. If you thrive more under structured expectations, academia would be better for you, for although industry has all the trappings of long-range strategies and career planning, the highly reactive environment means these are more honored in the breach. For one thing, reorganizations are common, and in the extreme case mergers (I have experienced two) can reset everything, for good or ill, and devour many months.

This is not to say that all is chaos—industry certainly favors a goal-directed personality, but with plenty of flexibility. On the other hand, flexibility is more the hallmark of academic research, where you will have the opportunity to follow wherever the science leads, once you are running your own shop. In industry, the flexibility is more of the conforming sort, since you won't be able to investigate every promising lead and change your research direction at will. In academia, diverging from the Specific Aims of a grant may be a problem when the time comes to renew, but the risk is yours, as is the reward. In industry, you can make the case for a new program of research, but the decision is management's and will be guided by business considerations. The "lone wolf" or "one-person band" may be increasingly rare in academia in an age of collaboration, but it is unheard of in industry, where being able to work in teams with specialized division of labor is essential. It should be apparent, as well, that mavericks and quirky personalities tend to do better in academia.

The pecking order in industry is deeper and more pyramidal than in academia, and you might end up languishing in a pay grade (or feel like you are), but there are usually plenty of opportunities for lateral moves and a variety of experiences—not to mention that it's easier to switch companies than colleges. In industry, one does need to be able to thrive in a hierarchy; you will always answer to someone, though the degree to which you are monitored will vary. By the same token, if your personality is such that climbing a management ladder and assuming steadily greater responsibility suits you, industry is built for that, and plenty of management training is on offer in larger companies. Learning to manage is much more hit-or-miss in academia; opportunities to lead large organizations are rare (and to manage them actively rather than by consensus, rarer still).

If your personality type is that of a risk-taker, biotechs and/or startups may fit you to a tee. These are the wild and wooly end of the industry spectrum, and the risks and rewards are well-known. You will work longer hours than in large pharma, and maybe even more than in academia. You will most likely share more in ownership, and learn entrepreneurial skills that will serve you well, once the bug has bitten. Bear in mind the very common pattern of faculty spinning off startups or otherwise participating in boards and the like, not to mention staking out intellectual property (shared with their university); thus, you may well be able to scratch this itch from the vantage of academia as well.

A final word about politics. Whether you are an enthusiastically political animal, or abhor this aspect of the human condition, you will encounter plenty of politics in both academia and industry. The flavors differ, to be sure. As a student you doubtless heard the clichés about tedious academic committees and underhanded deans, but you have probably had more exposure to the realities behind those stories than the corresponding ones about the dog-eat-dog corporate world. Company politics, I would hazard to say, are more transparent—the maneuvering more open and the motives more apparent. The results are often more life-altering, unbuffered by tenure and academic convention. Again, it is a matter of taste, but in my opinion the differences are overblown, for the simple reason that people are the same everywhere, in both environments governed by an underlying sense of fair play, but also occasional opportunism.

## Rule 5: Consider the Alternatives

As I've suggested, the choice you face is far more fine-grained than simply that between industry and academia. Industry is a spectrum, from large pharma to mature biotech to startup. By the same token, the academic side has at one extreme the research powerhouses, where you will be judged by volume of grants, and at the other the teaching institutions, which may not even have graduate departments. Unless you are very sure of yourself, you'd be well-advised to consider the full range, given the competition you may face.

Also, don't neglect other careers that may value your training. If you love the language, consider science journalism, either writing or editing—*Science* and *Nature* have large staffs, and you will often encounter them and representatives of other journals at the same scientific meetings you attend. The same is true of government agencies such as the NIH, NSA, DOE, and so forth, where grants administration is very actively tied to research trends and can be an entrée into the world of science policy. There are many more such positions when foundations, interest groups, and other private funding bodies are included. If you have a knack for business, many management consulting firms have scientific and technical consulting arms that value Ph.D.s and offer intensive training opportunities, and, though it may not be attractive at the moment, a career as a financial analyst specializing in biotech is yet another possibility.

## Rule 6: Consider the Timing

The current business environment cannot help but be among your considerations. Pharma has certainly been contributing to the unemployment rolls of late. Corporate strategies, which used to be very similar across the sector, have started to diverge, so that some companies are divesting bioinformatics at the same time that others are hiring computational types disproportionately as they place more of an emphasis on mathematical modeling, systems approaches, pharmacogenomics, drug repurposing, and the like. Overall, though, the industry trend has been to shrink R&D, and this may well continue through a round of consolidation, with several mega-mergers now under way. As noted above, mergers are times of upheaval, carrying both risk and opportunity, and usually a period in limbo as well. At the same time, it is worth bearing in mind that a corollary of downsizing is outsourcing, so that there may be new opportunities for startups and even individual consultants.

For much of the last decade, academia has also been in the doldrums, as NIH budgets have effectively contracted. As I write this, things are definitely looking up, with prospects for renewed funding of science and even near-term benefits to the NIH and NSA from the Obama stimulus package. Whether universities will respond proportionately with faculty hiring, given the losses in their endowment funds and cutbacks in salaries and discretionary spending, remains to be seen. There is a lot of slack to be taken up, and in particular a backlog of meritorious grant applications that are now being reconsidered. Nevertheless, on balance, an academic career has to be somewhat more promising today than a year ago, and a career in pharma rather less so, in the opinion of the author.

## Rule 7: Plan for the Long Term

Having noted the current situation in Rule 6, it's important also to say that a career decision should be made with the long haul in mind. The business cycle will eventually reverse itself, and while the business model may need to change irrevocably, the aging population alone dictates that healthcare will be an increasing global priority. Likewise, history shows that growth in government funding for science waxes and wanes, with a time constant somewhat longer than a decade. Trying to optimize a career decision based on current conditions is a bit like trying to time the stock market—you are sure to be overtaken by events.

One approach is to choose some reasonably long time frame, perhaps a decade, and ask yourself whether you'd be content to have lived through the average ups and downs you'd experience in a given job over that period. In academia, that would include a tenure decision (rate your chances), a lot of grant applications with mixed success at best, and maybe some great students and really significant scientific contributions. In pharma or large biotech, it would encompass a couple of promotions, your own group and maybe a department, at least one merger or other big disruption, and several rounds of layoffs. In small business, it might include a failed startup (or two, or three), an IPO if you're lucky, and a lucrative exit strategy or long-term growth if you're really lucky.

If you game these scenarios with various probabilities, and use your imagination, it just might become clear which ones you have no stomach for, and which ones really hold your interest.

## Rule 8: Keep Your Options Open

Job-hopping is much more prevalent now than in days of yore, and you should consider this in your scenarios. In industry, there is little stigma attached to changing employers, and if you can tolerate the relocation and/or want to see the world, it is a more or less standard way to advance your career by larger-than-usual increments. This stratagem is far from unknown in academia, but perhaps a bit trickier to execute, though of course it is de rigueur if you fail to get tenure.

Of greater interest is the question of moving between academia and industry. From the former to the latter is fairly easy, but the reverse is not as common, for a variety of reasons. Superstar academics in relevant areas are in great demand in industry, to which they are often exposed through consulting or scientific advisory boards. There are multiple examples of senior academics taking over major R&D organizations in industry, sometimes orders of magnitude larger than anything they managed in academia, and you might even consider this well-trod path as a career goal from the outset.

It is not impossible to return to academia from industry, particularly if you were already quite prominent when you left, but if you start your career in industry you may be at a disadvantage unless you go to great lengths to maintain an academic-style publication record and CV. Important exceptions would be if the work that you did in industry was particularly novel and/or high-profile, or if your business experience is valued in the post you seek. Examples of the latter might be faculty positions with a prominent management component (centers, institutes, core facilities, and the like), or an interface role back to industry, or perhaps a joint business school appointment.

## Rule 9: Be Analytic

Approach the decision with the analytic skills you've learned to apply to scientific questions. Gather data from all available sources and organize it systematically. When you interview, don't just impress, but get impressions; record everything down to your gut feelings. Do some bibliometric or even social network analyses of your potential colleagues. Check the industry newsletters and blogs, albeit with a grain of salt, to get a sense of the mood around R&D units (not to be confused with manufacturing, sales and marketing, or other divisions, which may have completely different cultures within the same company).

You might even try out some decision theoretic methodologies, such as decision matrices and Bayesian decision trees, or run simulations on the scenarios of Rule 7. I recommend taking a look at expected utility theory and prospect theory, for an interesting quantitative excursion. But honestly, these suggestions are just a more sophisticated informatics version of the classic advice to "make a list of pros and cons," which always makes one feel a little more in control.

## Rule 10: Be Honest with Yourself

Another homily: Now, if ever, is the time to be honest with yourself. Take a hard look at your qualifications, with as much objectivity as you can muster, and use these rules to decide where you would be best-suited and positioned for success. But even more importantly, deal with your emotional responses to industry and academia. If something is nagging at you, tease it out into the open, and try to decide if it is well-founded or not; if you can't decide, then you have to acknowledge it, and realize that it may not go away in the future either.

Finally, try to keep some perspective. Your career choice is important, but not irrevocable, and there are more consequential things in life. Don't let the decision process ruin what should be an exciting time for you.

Editorial

# Ten Simple Rules for Aspiring Scientists in a Low-Income Country

**Edgardo Moreno[1]*, José-María Gutiérrez[2]**

1 Programa de Investigación en Enfermedades Tropicales, Escuela de Medicina Veterinaria, Universidad Nacional, Heredia, Costa Rica, 2 Instituto Clodomiro Picado, Facultad de Microbiología, Universidad de Costa Rica, San José, Costa Rica

Being a scientist entails a common set of characteristics. Admiring nature and having concern for social issues; possessing a strong academic background, team work abilities, honesty, discipline, skepticism, communication skills, competitiveness, ability to accept and give criticism, and productive relationships are some of the most obvious traits that scientists should have. To be a scientist in a low-income country (LIC), however, requires a complementary set of qualities that are necessary to confront the drawbacks that work against the development of science. The failure of many young researchers to mature as professional scientists upon their return to their country from advanced training elsewhere, motivated us to propose these ten rules.

## Rule 1: Understand Your Country

Most LIC scientists want to live in their home country. Nevertheless, you must be realistic and prepared to face rudimentary laboratories, power cuts, poor water supply, deficient libraries, slow Internet, and scarce or non-existent national funds for supporting research, hiring personnel, and providing maintenance or equipment. You must understand that science is a minor component of the cultural environment of an LIC and that, for most people and many politicians, science is a curiosity performed in high-income countries [1]. Within this adverse scenario, you should establish broad and strong links with your community and country. This involves becoming interested in historical, social, and political issues. LIC researchers have to enjoy the idiosyncrasies of their country, and cultivate the desire to contribute to the scientific development of their homeland and to the well-being of its people. Do not endorse deep doubts about the possibilities of performing research. It can be done—but not alone. Try to join efforts with other investigators facing the same problems. Learn how they sidetrack difficulties, and incorporate yourself into a research team. If you are not able to find a group that fits your specific interest, then procure a group of researchers who, although investigating topics marginal to your own, are capable of understanding the relevance of your work. At the initial phases of your career, belonging to a creative scientific environment in which your knowledge and skills are appreciated is of major importance. Be part of a team before trying to lead one.

## Rule 2: Focus on Your Scientific Work

Your formal education has finished, but your scientific career is just beginning. Research should be your main professional activity. Consider that you may be the country's only specialist in a particular topic, but keep in mind that science is global. You are a small fish in a big pond and part of an international community. Grow within this global context. Concentrate on your work, and do not pay attention to flattering comments. Above all, keep away from activities that distract you from scientific endeavor, such as excessive administrative duties, and too many committees. Limit the number of meetings and attend only the relevant ones. Even though you are well prepared, modestly declare yourself as "ignorant" in topics that may distract you, and fight against excessive lecturing. However, participate in graduate programs and seminars. This is the right environment for the promotion of academic knowledge and skills.

## Rule 3: Be Wise When Selecting Your Research Topic

LICs face many problems that await creative solutions. Bizarre as it sounds, you can turn this into an advantage since these same problems constitute excellent sources for research and offer comparative advantages. Try to choose a topic that is not directly pursued by many or strong international research teams. At the beginning of your career, you cannot compete with them and your efforts may be frustrated. Identify the potential bottlenecks. Remember that in LICs research time runs slower and that good science is not so much related to the subject as to the answers you extract from your investigations. Frequently, local models become universal once a coherent story is built around them. Become an expert and, simultaneously, broaden your knowledge in collateral areas that may open new possibilities.

## Rule 4: Improve Your Communication Skills

English is the language of natural sciences, and you cannot avoid this fact. Consequently, you should be proficient in this language. The international scientific community is lenient about strong accents. However, the same community does not tolerate poor writing. Thus, writing skills are essential, since research begins with written proposals [2] and does not end until your results have been published [3]. You, more than native English speakers, must practice your oral presentations [4].

## Rule 5: Collaborate Locally and Internationally

Collaboration is essential for the advancement of science. Although this holds true for any researcher in the world [5], it is crucial for LIC investigators. Identify local groups who share your scientific

**Citation:** Moreno E, Gutiérrez J-M (2008) Ten Simple Rules for Aspiring Scientists in a Low-Income Country. PLoS Comput Biol 4(5): e1000024. doi:10.1371/journal.pcbi.1000024

**Published** May 30, 2008



**Funding:** No funding was received for the work presented.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: emoreno@medvet.una.ac.cr

interest, have equipment, or perform activities or techniques that are useful for your research. Keep in touch with your former tutor and colleagues and explore new collaborations abroad. Do not be shy about requesting help, and offer something that attracts the attention of your counterparts. Attend international meetings and present your work. Research is, in a way, a trade market of ideas, methods, and goods. Travel and visit research institutions. If some experiments cannot be carried out in your country, arrange to perform them abroad, or convince people to do them for you. There are international funds available for this purpose.

## Rule 6: Commit Yourself to the Education of Young Scientists

LIC researchers should participate in graduate training programs since this is the best way to build a strong scientific community. It is also a way to identify good young students and potential partners. Carefully choose the subjects for your students, pondering the possibilities of your research center, and be realistic about what they can achieve and the tasks you are imposing on them. Upgrade your students' education by sending them abroad for seminars and for learning specific methodologies (http://iscbsc.org/scs3/index.htm). There are international fellowships for this purpose (http://www.twas.org/). Be strict but generous with your students and colleagues, and, whenever possible, share your facilities and knowledge. Do not be self-centered. Promoting the success of others is also a way to promote your own success.

## Rule 7: Write Research Grants and Publish in International Journals

Scientific amateurism is common in LICs. Science is not a hobby but a professional activity that requires strong commitment. Inform yourself about local and international granting agencies, and apply for money [2]. There are international agencies and programs that provide grant and travel funds for LIC investigators (e.g., TWAS, IFS, EU, NIH, etc.). Although funds are limited, they will help you to build your scientific career. Incorporate yourself into international consortia; they may find your ideas and resources interesting. If you do not have access to essential publications, send requests to authors, editors, or colleagues abroad. Avoid publishing your results in magazines or low-quality journals, and instead submit your work to international journals. Do not overestimate or underestimate your work, be realistic when choosing a suitable journal [3], and, above all, do not be overly frustrated when grants or papers are rejected; instead, use the experience as a source of learning. Even though some reviewers may undervalue research performed in LICs, most of them pay more attention to the results and ideas than to nationalities [6].

## Rule 8: Develop Endurance When Confronting Difficulties

It is understandable that the limitations of performing research in LICs sometimes weaken your enthusiasm. Remain calm and try to identify the source of the problem; avoid complaining excessively in front of students, colleagues, or your partners abroad. A negative attitude is contagious, lowers your prestige, and has the tendency to attract unproductive people. Share your problems with other local scientists and confront them as a team. You should cultivate your abilities to find alternative solutions, as well as skills to improvise and to persuade people.

## Rule 9: Educate Yourself as a Professional Scientist

To be a specialist in an LIC is not enough. Be aware that the scientific community in an LIC is in short supply and lacks redundancy. In order to confront the drawbacks and deficiencies of the system, you must acquire a wide scientific knowledge, and become a well educated person in a broad sense. In addition to helping the quality of your research, this will give you the credentials to participate in political decisions related to science, to promote your ideas, and to spread scientific knowledge in your country. Acquaint yourself with local and international trends related to scientific performance and keep track of the major breakthroughs in science. Give talks and write about science whenever you consider it pertinent, but without diverting your attention too much from your main scientific duties.

## Rule 10: Appreciate Being a Scientist

As most scientists from high income countries and from LICs know, we are prone to facing economic difficulties at the beginning of our careers. Generally, salaries for scientists are comparatively low. Nevertheless, in time scientists can achieve a satisfying income; furthermore, there are compensations, especially if you become a successful scientist. A sense of achievement and contribution to your community, prestige, travel, meeting interesting people, and consulting opportunities are some of them, but nothing is more rewarding than the intellectual stimulation of science itself. This was your original motivation; nourish it with more and better science.

## Acknowledgments

The style for this article was inspired by the "Ten Simple Rules…" papers published by Philip E. Bourne in *PLoS Computational Biology*. We acknowledge the comments and revision of our colleagues and former students.

## References

1. Moreno E, Alveteg T (2002) Collaboration between Sweden and the Public Universities of Nicaragua. Swedish International Development Cooperation Agency (Sida, Evaluation 03/31). Stockholm. http://www.oecd.org/dataoecd/43/21/35213123.pdf.
2. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comput Biol 2: e12. doi:10.1371/journal.pcbi.0020012.
3. Bourne PE (2005) Ten simple rules for getting published. PLoS Comput Biol 1: e57. doi:10.1371/journal.pcbi.0010057.
4. Bourne PE (2007) Ten simple rules for making good oral presentations. PLoS Comput Biol 3: e77. doi:10.1371/journal.pcbi.0030077.
5. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3: e44. doi:10.1371/journal.pcbi.0030044.
6. Yousefi-Nooraie R, Shakiba B, Mortaz-Hejri S (2006) Country development and manuscript selection bias: a review of published studies. BMC Med Res Methodol 6: 37.

Editorial

# Ten Simple Rules for Selecting a Postdoctoral Position

Philip E. Bourne*, Iddo Friedberg

You are a PhD candidate and your thesis defense is already in sight. You have decided you would like to continue with a postdoctoral position rather than moving into industry as the next step in your career (that decision should be the subject of another "Ten Simple Rules"). Further, you already have ideas for the type of research you wish to pursue and perhaps some ideas for specific projects. Here are ten simple rules to help you make the best decisions on a research project and the laboratory in which to carry it out.

## Rule 1: Select a Position that Excites You

If you find the position boring, you will not do your best work—believe us, the salary will not be what motivates you, it will be the science. Discuss the position fully with your proposed mentor, review the literature on the proposed project, and discuss it with others to get a balanced view. Try and evaluate what will be published during the process of your research. Being scooped during a postdoc can be a big setback. Just because the mentor is excited about the project does not mean you that will be six months into it.

## Rule 2: Select a Laboratory That Suits Your Work and Lifestyle

If at all possible, visit the laboratory before making a decision. Laboratories vary widely in scope and size. Think about how you like to work—as part of a team, individually, with little supervision, with significant supervision (remembering that this is part of your training where you are supposed to be becoming independent), etc. Talk to other graduate students and postdoctoral fellows in the laboratory and determine the work style of the laboratory. Also, your best work is going to be done when you are happiest with the rest of your life. Does the location of the laboratory and the surrounding environment satisfy your nonwork interests?

## Rule 3: Select a Laboratory and a Project That Develop New Skills

Maximizing your versatility increases your marketability. Balance this against the need to ultimately be recognized for a particular set of contributions. Avoid strictly continuing the work you did in graduate school. A postdoctoral position is an extension of your graduate training; maximize your gain in knowledge and experience. Think very carefully before extending your graduate work into a postdoc in the same laboratory where you are now—to some professionals this raises a red flag when they look at your resume. Almost never does it maximize your gain of knowledge and experience, but that can be offset by rapid and important publications.

## Rule 4: Have a Backup Plan

Do not be afraid to take risks, although keep in mind that pursuing a risky project does not mean it should be unrealistic: carefully research and plan your project. Even then, the most researched, well-thought-out, and well-planned project may fizzle; research is like that. Then what? Do you have a backup plan? Consider working on at least two projects. One to which you devote most of your time and energy and the second as a fallback. The second project should be more of the "bread and butter" type, guaranteed to generate good (if not exciting) results no matter what happens. This contradicts *Rule 1,* but that is allowed for a backup plan. For as we see in *Rule 5,* you need tangible outcomes.

## Rule 5: Choose a Project with Tangible Outcomes That Match Your Career Goals

For a future in academia, the most tangible outcomes are publications, followed by more publications. Does the laboratory you are entering have a track record in producing high-quality publications? Is your future mentor well-respected and recognized by the community? Talk to postdocs who have left the laboratory and find out. If the mentor is young, does s/he have the promise of providing those outcomes? Strive to have at least one quality publication per year.

## Rule 6: Negotiate First Authorship before You Start

The average number of authors on a paper has continued to rise over the years: a sign that science continues to become more collaborative. This is good for science, but how does it impact your career prospects? Think of it this way. If you are not the first author on a paper, your contribution is viewed as $1/n$ where $n$ is the number of authors. Journals such as this one try to document each author's contributions; this is a relatively new concept, and few people pay any attention to it. Have an understanding with your mentor on your likelihood of first authorship before you start a project. It is best to tackle this problem early during the interview process and to achieve an

**Citation:** Bourne PE, Friedberg I (2006) Ten simple rules for selecting a postdoctoral position. PLoS Comput Biol 2(11): e121. doi:10.1371/journal.pcbi.0020121



Philip E. Bourne is a professor in the Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America, and is Editor-in-Chief of *PLoS Computational Biology*. Iddo Friedberg is a research assistant in the Bioinformatics and Systems Biology program at the Burnham Institute for Medical Research, La Jolla, California, United States of America.

* To whom correspondence should be addressed. E-mail: bourne@sdsc.edu

understanding; this prevents conflicts and disappointments later on. Don't be shy about speaking frankly on this issue. This is particularly important when you are joining an ongoing study.

## Rule 7: The Time in a Postdoctoral Fellowship Should Be Finite

Mentors favor postdocs second only to students. Why? Postdocs are second only to students in providing a talented labor pool for the least possible cost. If you are good, your mentor may want you to postdoc for a long period. Three years in any postdoc is probably enough. Three years often corresponds to the length of a grant that pays the postdoctoral fellowship, so the grant may define the duration. Definitely find out about the source and duration of funding before accepting a position. Be very wary about accepting one-year appointments. Be aware that the length of a postdoc will likely be governed by the prevailing job market. When the job market is good, assistant professorships and suitable positions in industry will mean you can transition early to the next stage of your career. Since the job market even a year out is unpredictable, having at least the option of a three-year postdoc fellowship is desirable.

## Rule 8: Evaluate the Growth Path

Many independent researchers continue the research they started during their postdoc well into their first years as assistant professors, and they may continue the same line of work in industry, too. When researching the field you are about to enter, consider how much has been done already, how much you can contribute in your postdoc, and whether you could take it with you after your postdoc. This should be discussed with your mentor as part of an ongoing open dialog, since in the future you may be competing against your mentor. A good mentor will understand, as should you, that your horizon is independence—your own future lab, as a group leader, etc.

## Rule 9: Strive to Get Your Own Money

The ease of getting a postdoc is correlated with the amount of independent research monies available. When grants are hard to get, so are postdocs. Entering a position with your own financing gives you a level of independence and an important extra line on your resume. This requires forward thinking, since most sources of funding come from a joint application with the person who will mentor you as a postdoc. Few graduate students think about applying for postdoctoral fellowships in a timely way. Even if you do not apply for funding early, it remains an attractive option, even after your postdoc has started with a different funding source. Choosing one to two potential mentors and writing a grant at least a year before you will graduate is recommended.

## Rule 10: Learn to Recognize Opportunities

New areas of science emerge and become hot very quickly. Getting involved in an area early on has advantages, since you will be more easily recognized. Consider a laboratory and mentor that have a track record in pioneering new areas or at least the promise to do so. ■


## Acknowledgments

The authors would like to thank Mickey Kosloff for helpful discussions.

Editorial

# Ten Simple Rules for Getting Grants

Philip E. Bourne*, Leo M. Chalupa

This piece follows an earlier Editorial, "Ten Simple Rules for Getting Published" [1], which has generated significant interest, is well read, and continues to generate a variety of positive comments. That Editorial was aimed at students in the early stages of a life of scientific paper writing. This interest has prompted us to try to help scientists in making the next academic career step—becoming a young principal investigator. Leo Chalupa has joined us in putting together ten simple rules for getting grants, based on our many collective years of writing both successful and unsuccessful grants. While our grant writing efforts have been aimed mainly at United States government funding agencies, we believe the rules presented here are generic, transcending funding institutions and national boundaries.

At the present time, US funding is frequently below 10% for a given grant program. Today, more than ever, we need all the help we can get in writing successful grant proposals. We hope you find these rules useful in reaching your research career goals.

## Rule 1: Be Novel, but Not Too Novel

Good science begins with new and fresh ideas. The grant writing process should be a pleasure (no, we are not kidding), for it allows you to articulate those ideas to peers who have to read your grants but not necessarily your papers. Look at grant writing as an opportunity to have an impact. Feel passionate about what you are writing—if you are not passionate about the work, it is probably not a good grant and is unlikely to get funded. "Me-too" science will not get funded when funding levels are low. On the other hand, science that is too speculative will not be supported either, particularly when funds are tight—sad but true.

## Rule 2: Include the Appropriate Background and Preliminary Data as Required

You need to convince reviewers that the work you propose needs to be done and that you are the best person to do it. Different granting programs require differing amounts of preliminary data. For certain programs, it can be said that the work must be essentially done before the grant is awarded, and that the funds are then used for the next phase of the research program. There is some truth in this. So where appropriate, do provide some tantalizing preliminary result, making sure to tell the reviewers what these results imply with respect to the specific aims of your proposal. In formulating the motivation for your proposal, make sure to cite all relevant work—there is nothing worse than not appropriately citing the work of a reviewer! Finally, convince the reviewer that you have the technical and scientific background to perform the work as proposed.

## Rule 3: Find the Appropriate Funding Mechanism, Read the Associated Request for Applications Very Carefully, and Respond Specifically to the Request

Most funding organizations have specific staff to assist in finding funding opportunities, and most funding agencies have components of their Web sites designed to help investigators find the appropriate programs. Remember, programs want to give away money—the jobs of the program's staff depend on it. The program staff can help you identify the best opportunities. If your grant does not fit a particular program, save your time and energy, and apply elsewhere, where there is a better programmatic fit.

## Rule 4: Follow the Guidelines for Submission Very Carefully and Comply

Many funding bodies will immediately triage grants that do not comply with the guidelines—it saves the program time and money. This extends to all the onerous supporting material—budget justification, bibliographies, etc. Get them right and keep them updated for future applications. Even if it goes to review, an inappropriately formulated application may aggravate the reviewers, and will have a negative impact even if the science is sound. Length and format are the most frequent offenders.

## Rule 5: Obey the Three Cs—Concise, Clear, and Complete

The grant does not have to fill the allotted page count. Your goal should be to provide a complete reckoning of what is to be done, as briefly as possible. Do not rely on supplements (which may not be allowed) or on Web sites (review may be actively discouraged since it has the potential to compromise anonymity). Specify the scope up-front and make sure it is realistic with respect to the funds requested. A common temptation for inexperienced grant writers is to propose to do too much. Such applications are usually judged as overly ambitious and consequently poorly rated.

## Rule 6: Remember, Reviewers Are People, Too

Typically, reviewers will have a large number of grants to review in a short period. They will easily lose concentration and miss key points of your proposal if these are buried in an overly lengthy or difficult-to-read document. Also, more than likely, not all the reviewers will be experts in your

**Citation:** Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comput Biol 2(2): e12.



Philip E. Bourne is a professor in the Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America, and is Editor-in-Chief of *PLoS Computational Biology*. Leo M. Chalupa is a professor and chair in the Section of Neurobiology, Physiology, and Behavior, University of California Davis, Davis, California, United States of America.

**DOI:** 10.1371/journal.pcbi.0020012

* To whom correspondence should be addressed. E-mail: bourne@sdsc.edu

discipline. It is a skill to capture the interest of experts and nonexperts alike. Develop that skill. Unlike a paper, a grant provides more opportunity to apply literary skills. Historical perspectives, human interest, and humor can all be used judiciously in grants to good effect. Use formatting tricks (without disobeying rule 4), for example, underlining, bolding, etc., and restate your key points as appropriate. Each section can start with a summary of the key points.

### Rule 7: Timing and Internal Review Are Important

Give yourself the appropriate lead time. We all have different approaches to deadlines. Ideally, you should complete a draft, leave sufficient time to get feedback from colleagues, and then look at the grant again yourself with a fresh eye. Having a spectrum of scientific colleagues who are similar to the likely reviewer pool critique your grant is very valuable.

### Rule 8: Know Your Grant Administrator at the Institution Funding Your Grant

At the end of the day, this person is your best advocate. How well you understand each other can make a difference. Many grant administrators have some measure (limited to complete) discretionary control over what they fund. The more they know and understand you and your work, the better your chances of success. Do not rely just on E-mail to get to know the grant administrator. Do not be intimidated. Talk to them on the telephone and at meetings where possible—they want to help.

### Rule 9: Become a Grant Reviewer Early in Your Career

Being on review panels will help you write better grants. Understanding why grants get triaged before complete review, how a panel reacts to a grant, what the discretionary role of program officers is, and what the role of oversight councils is provide valuable lessons for writing successful grants of your own and for giving others advice about this process.

### Rule 10: Accept Rejection and Deal with It Appropriately

Rejection is inevitable, even for very good grants when funding levels are low. Learn to live with rejection and to respond appropriately. Do not be defensive; address each criticism head on and respond with facts and not emotional arguments. When resubmission is necessary, make it very clear to the reviewer that you understand what was wrong the first time. Indicate precisely how you have fixed the problems. In the resubmitted application, never argue with the validity of the prior review. If the grant was close to being funded the first time around, remind the reviewers of that fact by including the previous score if appropriate, and make it crystal clear why this version is much improved.

There are no previously unrevealed secrets to grant writing presented here. Rather, it is a concise picture intended to help our early career readers take the next step. If you feel like you need more detail, take a look at Kraicer's article [2]. Good luck on getting those grants. ■


**References**

1. Bourne PE (2005) Ten simple rules for getting published. PLoS Comput Biol 1: DOI: 10.1371/journal.pcbi.0010057
2. Kraicer J (1997) The art of grantmanship. Strasbourg: Human Frontier Science Program. Available: http://www.hfsp.org/how/ArtOfGrants.htm. Accessed 19 January 2006.

PLOS | COMPUTATIONAL BIOLOGY

EDUCATION

# Ten simple rules for helping newcomers become contributors to open projects

**Dan Sholler[1☯], Igor Steinmacher[2☯], Denae Ford[3☯], Mara Averick[4☯], Mike Hoye[5☯], Greg Wilson[6☯]***

**1** Berkeley Institute for Data Science, University of California Berkeley, Berkeley, California, United States of America, **2** School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, United States of America, **3** Department of Computer Science, North Carolina State University, Raleigh, North Carolina, United States of America, **4** RStudio, Inc., Boston, Massachusetts, United States of America, **5** Mozilla Corporation, Toronto, Ontario, Canada, **6** RStudio, Inc., Toronto, Ontario, Canada

☯ These authors contributed equally to this work.
* greg.wilson@rstudio.com

Check for updates



**Citation:** Sholler D, Steinmacher I, Ford D, Averick M, Hoye M, Wilson G (2019) Ten simple rules for helping newcomers become contributors to open projects. PLoS Comput Biol 15(9): e1007296. https://doi.org/10.1371/journal.pcbi.1007296

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** September 12, 2019



**Funding:** The authors received no specific funding for this study.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

To survive and thrive, a community must attract new members, retain them, and help them be productive [1]. As openness becomes the norm in research, software development, and education, knowing how to do this has become an essential skill for principal investigators and community managers alike. A growing body of knowledge in sociology, anthropology, education, and software engineering can guide decisions about how to facilitate this.

What exactly do we mean by "community"? In the case of open source and open science, the most usual meaning is a "community of practice." As defined by Lave and Wenger [2, 3], groups as diverse as knitting circles, oncology researchers, and web designers share three key characteristics:

1. Participants have a common product or purpose that they work on or toward.
2. They are mutually engaged, i.e., they assist and mentor each other.
3. They develop shared resources and domain knowledge.

Brown [4] specializes this to define a "community of effort" as

> . . .a community formed in pursuit of a common goal. The goal can be definite or indefinite in time, and may not be clearly defined, but it is something that (generally speaking) the community is aligned on.

People working to preserve coral reefs in the face of global climate change are an example of such a community. No central organization coordinates their work, but the scientists who study coral reefs, the environmentalists who work to protect them, and the citizens who support them financially and politically are aware of each other's efforts, collaborate in ad hoc ways, and are conscious of contributing toward a shared purpose.

Open-source software projects are also communities of effort. E.g., the Mozilla Firefox [5] community includes a mix of paid professionals, highly involved volunteers, and occasional contributors who not only create software, documentation, and tutorials but also organize events, answer questions in online forums, mentor newcomers, and advocate for open standards.

Every community of effort has unique features, but they have enough in common to profit from one another's experience. The 10 rules laid out below are based on studies of such communities and on the authors' experience as members, leaders, and observers. Our focus is on small and medium-sized projects, i.e., ones that have a handful of to a few hundred participants and are a few months to a few years old but may not (yet) have any formal legal standing, such as incorporation as a nonprofit.

## Rule 1: Be welcoming

Karl Fogel wrote [6], "If a project doesn't make a good first impression, newcomers may wait a long time before giving it a second chance". Other authors have empirically confirmed the importance of kind and polite social environments in open-source projects [7–9]. Therefore, projects should not just say that they welcome new members: they should make a proactive effort to foster positive feelings in them. One way to do this is to post a welcome message on the project's social media pages, Slack channels, forums, or email lists. Projects might also consider maintaining a dedicated "Welcome" channel or list, where a project lead or community manager writes a short post asking newcomers to introduce themselves.

Other ways to be welcoming include offering assistance in finding ways to make an initial contribution, directing the newcomer to project members who have a similar background or skillset so as to demonstrate fit to the newcomer, and pointing the newcomer to essential project resources (e.g., the contribution guidelines). It also helps to clearly identify work items they can start with; a growing number of projects explicitly tag bugs or issues as "suitable for newcomers" and ask established members not to fix them in order to ensure there are suitable places for new arrivals to start work.

Projects can further designate one or two members to serve as a point of contact for each newcomer. Doing this may reduce the newcomer's hesitancy to ask questions, particularly when they are told from the outset that there are no dumb questions in the community.

## Rule 2: Help potential contributors evaluate if the project is a good fit

People could contribute to many different projects; the first and most important step in being welcoming is to help them determine whether your project is a good fit for their interests and abilities. Their decision to contribute can be related to reputation or external needs but also to a desire to learn or give back to the community. In all of these cases, the more you help newcomers understand whether this is the right project for them, the more quickly they will either start contributing or look elsewhere.

To do this, the project should explicitly state what the different types of skills required are. This information should be easily accessible and guide new members to the tasks they may handle. LibreOffice, e.g., provides a way for developers to filter available tasks by required skills and difficulty [10].

The project should also help developers evaluate their skills, since "basic Python skills" means very different things to different people. Tools like My GitHub Resume [11] and Visual Resume [12] that aggregate information from previous contributions can help with this assessment, while the now-defunct OpenHatch project [13] aggregated entry-level issues from a variety of open-source projects and classified them according to language and other required skills to provide a one-stop portal for finding appropriate projects.

## Rule 3: Make governance explicit

Raymond's "The Cathedral and the Bazaar" [14] described an egalitarian world in which everyone could contribute equally to open projects. Two decades later, we can see how unequal and unwelcoming the supposedly egalitarian "bazaar" of open source can be if authority lies with those willing to shout loudest and longest. As Bezroukov pointed out [15], Raymond ignored the realities of how power arises, becomes concentrated in a few hands, and is then used to perpetuate itself.

Bezroukov's criticism drew on Freeman's influential essay "The Tyranny of Structurelessness" [16], which explained how an apparent lack of structure in organizations " . . .too often disguised an informal, unacknowledged and unaccountable leadership that was all the more pernicious because its very existence was denied". The solution is to make a project's governance explicit so that people know who makes which decisions.

Large, well-established projects that incorporate as nonprofits are required to promulgate bylaws, such as those for the Python Software Foundation [17]. What smaller projects should do is less well-documented but generally falls under one of three headings [6]. The first is a "benevolent dictator" (often the project founder), who the community agrees has final say on important issues. This model is common in young or small projects but is brittle and inevitably fosters the emergence of unofficial (and hence unaccountable) de facto leaders in specific areas.

The second model formalizes a consensus-building process in which the whole community can take part. One example is Martha's Rules [18], under which anyone can put forward proposals, but those proposals are only adopted once it is clear that most people are not strongly opposed. The third model is based on elected representation. In the Carpentries [19], e.g., the electorate includes anyone who has

- completed instructor certification in the preceding year;
- completed certification in the last two years and taught at least one workshop;
- been certified for more than two years and has taught at least twice in that time; or
- made a significant contribution to lesson development, infrastructure, or other activities as determined by the Executive Council.

Decisions are then made by those elected, though they may decide or be required to take some matters to a referendum vote.

More complex models are possible [20], but the most important thing is to decide on the rules well in advance of contentious issues emerging, since tempers may already be running hot by the time this point is reached.

## Rule 4: Keep knowledge up to date and findable

When starting to contribute to a project, newcomers must orient themselves in an unfamiliar landscape [21]. It is therefore important to make sure that all necessary information is both accessible and findable. A single project may use wikis, files in GitHub, shared Google Docs, old tweets or Slack messages, and email archives; keeping information about a specific topic in a single place and clearly defining the purpose of each communication medium saves newcomers from having to navigate multiple unfamiliar data sources to find what they need. Doing this makes newcomers more confident and oriented [22].

At the same time, outdated documentation may lead newcomers to a wrong understanding of the project, which is also demotivating. While it may be hard to keep material up to date, community members should at least remove or clearly mark outdated information. Signaling the absence or staleness of material can save newcomers time and also suggest opportunities for them to make contributions that they themselves would find useful.

One special case of this rule is to provide "how to contribute" guidelines in easy-to-find, readily available places. Many projects follow GitHub's recommendation for placing such information in a CONTRIBUTING.md file [23]. Other projects, such as the Apache Open Office Suite and rOpenSci, provide newcomer manuals and learning modules accessed through a web interface [24, 25]. Still others take a more interactive approach; e.g., the GNOME project's Newcomers' Guide [26] walks potential contributors through the contribution pipeline: choosing a project, acquiring and installing the necessary computing tools, finding problems or choosing issues to work on, submitting changes, and following up on feedback.

Such guidelines do more than just describe how to contribute. First, their mere existence can ease newcomers' hesitation about whether or not their work is sufficient and suitable for the project. Second, they provide a centralized, well-organized description of resources that a newcomer can consult while learning to navigate the project's technical and social environments [27]. Guidelines also acclimate newcomers to the norms of work and communication, particularly when items such as necessary computing tools and codes of conduct are foregrounded.

## Rule 5: Have and enforce a code of conduct

Community leaders should model the behaviors they want to encourage, but that by itself is not enough: experience shows that communities must also make norms about acceptable behavior explicit. This helps ensure that everyone, not just newcomers, will find the environment healthy and welcoming. It also sends a clear signal that the community actually has standards: many potential contributors will be painfully familiar with communities that don't and are more likely to give yours a try if they believe it is not just another troll-infested chat room. Being explicit also makes the project more accessible to people from differing cultural backgrounds because it helps them understand how expectations may differ from what they are used to.

A popular way to make norms explicit is to adopt a code of conduct. Research on these is still in its infancy [28], but many projects such as rOpenSci [29], NumPy [30], and Project Jupyter [31] have adopted the Contributor Covenant [32] or used other frameworks such as SciPy's Code of Conduct [33].

A code of conduct is only useful if there is a clear reporting mechanism that community members trust and if it is enforced [34]. Projects should designate an independent party (i.e., an individual not employed by or otherwise closely connected to the project) to receive and review reports. An independent party offers a degree of objectivity and can help protect reporters from hesitating to raise issues concerning project leaders out of fear of retribution or damage to their reputation. When possible, the independent party should be part of a more extensive code of conduct committee made up of several people with varied characteristics (e.g., gender identity, race, ethnicity, roles in the community). Any member of the committee implicated in the incident should recuse themselves from reviewing the violations.

Project leaders should also develop and publicize enforcement mechanisms, which may range from verbal or written warnings, limits on access to project communication avenues (e.g., Slack channels or mailing lists), or suspension or expulsion from contributing to the project. When safe for the reporter, project leaders should also publicize enforcement decisions: if this is not done, the community may come to believe that the code is meaningless.

## Rule 6: Develop forms of legitimate peripheral participation

A core concept in the theory of communities of practice is that of legitimate peripheral participation (LPP) [2, 3]. Newcomers become members of a community by participating in simple, low-risk tasks that further the goals of the community. Through these peripheral activities, newcomers become acquainted with the community's tasks, vocabulary, and governance so that they can ease into the project.

In communities such as GitHub, core activities such as committing code and submitting pull requests can be socially daunting for newcomers [35]. One way to encourage LPP in this case is to encourage newcomers to submit issues to a repository when they notice a bug or to join the dialog on recently submitted pull requests or issues. Another way is to have newcomers help with documentation, particularly with translation and localization, and a third (mentioned in Rule 3) is to mark some issues as suitable for newcomers.

Building multiple ways of participating in a community demonstrates the variety of approaches newcomers can take to join the community. This further demonstrates that there is not just one way to make technical contributions. E.g., the main form of interaction in the community on Stack Overflow is to ask a question and post an answer, but engaging in that type of interaction can present barriers for some users, including an intimidating community size and fear of negative feedback [36]. Thus, it is important to provide additional forms of participation. On Stack Overflow, this is demonstrated through the ability to edit questions and answer without the restriction of reputation points. Developing a pathway to participation can decrease the presence of barriers. In studying the evolution of how content is formed in these communities [37], newcomers can better understand the norms of a community and the best way to contribute [38].

## Rule 7: Make it easy for newcomers to get started

One way to facilitate LPP is to make it easy for newcomers to get set up so that they can start work on contributions. Getting set up to work on a project—going from "I want to help" to "I'm able to help" to "I'm helping"—is often someone's first experience as a community participant. Any complexity or confusion at this point is therefore a significant barrier to participation [39]. By treating the process of getting involved with the same care and attention you give to the product itself, you're making it clear that you value those contributors' time and effort and forestalling reactions like this [40]:

> I am still trying to build, because many errors occurred. . . I was expecting to move forward, because so far I did not have time to look at the source code. . . It is frustrating.

This work does not just benefit newcomers; it also helps retention of existing intermittent contributors, and the same work that makes your project more accessible to new contributors today will do the same for future you. Wheelchair ramps and the buttons that open heavy doors are not just used by those in wheelchairs: they are just as helpful to people with strollers or one too many bags of groceries. None of us are ever more than a sprained ankle away from desperately wanting that wheelchair ramp to be there. In that same vein, a drive failure will someday force you to download a gigabyte of data and reinstall some software, inevitably at the least convenient moment imaginable. There is therefore a lot to be gained from automating as much of your setup process you can and thoroughly documenting whatever you cannot.

## Rule 8: Use opportunities for in-person interaction—With care

Open-source software projects often rely heavily on remote workers communicating via text, audio, and video. Research on face-to-face and audio/video-mediated communication is mixed with regard to their comparative effectiveness [41–43] but demonstrates that each form has benefits and drawbacks. In-person interaction is valuable for uninterrupted, synchronous dialog and helps to establish mutual understanding in a streamlined way [44]. Projects can therefore benefit from engaging newcomers in in-person interaction from time to time.

According to Huppenkothen and colleagues [45], newcomers may particularly benefit from events that

> ". . .combine structured periods focused on pedagogy (often with an emphasis on statistical and computational techniques) and less structured periods devoted to hacks and creative projects, with the goal of encouraging collaboration and learning among people at various stages of their career."

Combining newcomer-friendly events and activities with larger gatherings such as conferences also amortizes participants' financial costs and travel time.

However, potential contributors might shy away from the project if they are introverted, suffer from social anxiety, or have had bad experiences in the past in face-to-face settings. A code of conduct helps allay these concerns, but some newcomers may still feel uncomfortable in group settings. In this case, not going to a meetup may leave them feeling less a part of the community.

Face-to-face communication also involves forms of information exchange that are not easily captured and archived for all project members to see. E.g., collocated project members might hash out ideas on whiteboards, by scribbling notes, or through informal chats. Even when transcribing and/or taking photos of these is possible, important contextual information may be lost [46]. Decisions and changes may seem to come out of nowhere when evaluated by a nonattendee, so project leads should develop universally accessible ways to communicate and explain the results of in-person activities.

## Rule 9: Acknowledge all contributions

People in open source sometimes joke that a programmer is someone who will do something for a laptop sticker that they would not do for a hundred dollars. The kernel of truth in this joke is that gratitude and recognition are the most powerful tools community builders have. It is therefore crucial to acknowledge newcomers' contributions and thank them for their work. Every hour that someone has given your project may be an hour taken away from their personal life or their official employment; recognize that fact and make it clear that while more hours would be welcome, you do not expect them to make unsustainable sacrifices.

To ensure completeness and fairness, every project should adopt and publicize guidelines describing what constitutes a contribution, how contributions will be acknowledged, and how they will be used. Who can use the data collected by the project for what purposes, and what attribution do they have to give? How must they acknowledge the project and/or its contributors? Who holds the copyright on contributed material? Most projects now place this information in files called LICENSE.md and CITATION.md and place a brief, readable summary in plain language in onboarding materials.

## Rule 10: Follow up on both success and failure

Once someone has carried their first contribution over the line, you and they are likely to have a better sense of what they have to offer and how the project can help them. Helping newcomers find the next problem they might want to work on or pointing them at the next thing they might enjoy reading is both helpful and supportive. In particular, encouraging them to help the next wave of newcomers is both a good way to recognize what they have learned and an effective way to pass it on.

Mentoring programs are a popular way to do this. However, their effectiveness appears mixed. [47] found that ". . .developers receiving deliberate onboarding support through mentoring were more active at an earlier stage than developers entering projects through conventional means". In contrast, [48] found that

> ". . .developers who join an organization through these programs are half as likely to transition into long-term community members than developers who do not use these programs. . . although developers who do succeed through these programs find them valuable."

One explanation for this disparity is that people become members of open projects for different reasons and hence respond to things like mentoring programs in different ways. E.g., Barcomb and colleagues identified four types of episodic or intermittent contributors to open-source projects [49], while Mäenpää and colleagues looked at how to reconcile the competing yet complementary needs of stakeholders in hybrid open/commercial projects [50]. More research is needed, but as openness becomes the norm in research, doing it well becomes a core skill for every researcher.

When they can, projects should also try to follow up on their failures. Why did potential contributors not become community members? Did they realize that the project wasn't a good fit (in which case, the overview may need an overhaul)? Was it too difficult to find a starting point or to get set up to start work (in which case information may need to be consolidated, tagged, filled in, or updated)? Or did they feel uncomfortable or undervalued (in which case the community may need to have a more difficult conversation)? The conversations with individuals should in most cases be confidential, but making the conclusions and corrective actions public is the best possible way to signal that you are serious about building the best community you can.

## Martha's rules

1. Anyone may put forward a proposal up to 24 hours before a meeting. Proposals must include a one-line summary, a brief description, any required background information, and a discussion of pros and cons (including alternatives).
2. Once a person has sponsored a proposal, they are responsible for it; the group may not discuss or vote on the issue unless the sponsor is present.
3. After the sponsor presents the proposal, a "sense" vote is taken prior to any discussion in which people indicate whether they like the proposal, can live with it, or are uncomfortable with it.
4. If all or most of the group likes or can live with the proposal, it moves to a formal vote with no further discussion.
5. If most of the group is uncomfortable with the proposal, it is postponed for further rework by the sponsor.
6. If some members are uncomfortable, they can briefly state their objections. After 10 minutes of moderated discussion, the facilitator calls for a yes-or-no vote on adoption. If a majority vote "yes", the proposal is implemented. Otherwise, the proposal is returned to the sponsor for further work.

## Acknowledgments

The authors are grateful to everyone who gave feedback on early versions of this paper, and particularly to Erin Robinson for her detailed, knowledgeable, and helpful critique.

## References

1. Qureshi I, Fang Y. Socialization in open source software projects: A growth mixture modeling approach. Organizational Research Methods. 2011; 14(1):208–238.
2. Lave J, Wenger E. Situated Learning: Legitimate Peripheral Participation. Cambridge, UK: Cambridge University Press; 1991.
3. Wenger E. Communities of Practice: Learning, Meaning, and Identity. Cambridge, UK: Cambridge University Press; 1999.
4. Brown CT. Sustaining open source: thinking about communities of effort. 2019 Mar 2 [cited 2019 Mar 21] In: Living in an Ivory Basement: Stochastic thoughts on science, testing, and programming [Internet]. http://ivory.idyll.org/blog/2019-communities-of-effort.html.
5. Mozilla Foundation. Mozilla Firefox [cited 2019 Mar 21]. https://www.mozilla.org/en-US/.
6. Fogel K. Producing Open Source Software: How to Run a Successful Free Software Project. Sebastopol, CA: O'Reilly Media; 2005.
7. Singh V. Newcomer integration and learning in technical support communities for open source software. In: Proceedings of the 17th ACM International Conference on Supporting Group Work—GROUP'12. New York City, NY: ACM Press; 2012. p. 65–74.
8. Steinmacher I, Wiese I, Chaves AP, Gerosa MA. Why do newcomers abandon open source software projects? In: 2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE'13). Institute of Electrical and Electronics Engineers (IEEE); 2013. p. 25–32.
9. Steinmacher I, Pinto G, Wiese IS, Gerosa MA. Almost There: A Study on Quasi-Contributors in Open-Source Software Projects. In: Proceedings of the 40th International Conference on Software Engineering (ICSE'18). New York City, NY: ACM Press; 2018. p. 256–266.
10. The Document Foundation. LibreOffice Easy Hacks by Required Skill; Accessed March 21, 2019. https://wiki.documentfoundation.org/Development/EasyHacks/by_Required_Skill.
11. Coallier D. My GitHub Resume; Accessed March 21, 2019. https://resume.github.io/.
12. Sarma A, Chen X, Kuttal S, Dabbish L, Wang Z. Hiring in the Global Stage: Profiles of Online Contributions. In: 2016 IEEE 11th International Conference on Global Software Engineering. ICGSE 2016. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE); 2016. p. 1–10.
13. OpenHatch [Internet]. OpenHatch [cited 2019 Mar 27] 2019. http://openhatch.org/.
14. Raymond ES. The Cathedral and the Bazaar. Sebastopol, CA: O'Reilly & Associates; 2001.
15. Bezroukov N. A Second Look at the Cathedral and the Bazaar. First Monday. 1999; 4(12) [cited 2019 Mar 27]. https://firstmonday.org/article/view/708/618.
16. Freeman J. The Tyranny of Structurelessness. The Second Wave. 1972; 2(1): 20.
17. Python Software Foundation Bylaws [Internet]. Python Software Foundation [cited 2019 Feb 16]. https://www.python.org/psf/bylaws/.
18. Minahan A. Martha's Rules. Affilia. 1986; 1(2):53–56. https://doi.org/10.1177/088610998600100206
19. The Carpentries Bylaws [Internet]. The Carpentries [cited 2019 Feb 16]. https://docs.carpentries.org/topic_folders/governance/index.html.
20. The Apache Software Foundation: How It Works [Internet]. The Apache Software Foundation [cited 2019 Mar 27]. https://www-us.apache.org/foundation/how-it-works.html.
21. Dagenais B, Ossher H, Bellamy RKE, Robillard MP, de Vries JP. Moving Into a New Software Project Landscape. In: Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering (ICSE'10). New York City, NY: ACM Press; 2010. p. 275–284.
22. Steinmacher I, Conte TU, Treude C, Gerosa MA. Overcoming open source project entry barriers with a portal for newcomers. In: Proceedings of the 38th International Conference on Software Engineering (ICSE'16). New York City, NY: ACM Press; 2016. p. 273–284
23. Setting Guidelines for Repository Contributors [Internet]. GitHub [cited 2019 Feb 16]. https://help.github.com/articles/setting-guidelines-for-repository-contributors/.
24. Apache Software Foundation. Introduction to Contributing to Apache OpenOffice [Internet]. The Apache Software Foundation [cited 2019 Feb 16]. https://openoffice.apache.org/orientation/intro-contributing.html.

25. rOpenSci Packages: Development, Maintenance, and Peer Review [Internet]. rOpenSci [cited 2019 Feb 16]. https://ropensci.github.io/dev_guide/.

26. GNOME Newcomers' Guide [Internet]. GNOME [cited 2019 Feb 16]. https://wiki.gnome.org/Newcomers.

27. Zanatta AL, Steinmacher I, Machado LS, de Souza CRB, Prikladnicki R. Barriers Faced by Newcomers to Software-Crowdsourcing Projects. IEEE Software. 2017; 34(2):37–43. https://doi.org/10.1109/ms.2017.32

28. Tourani P, Adams B, Serebrenik A. Code of Conduct in Open Source Projects. In: 2017 24th International Conference on Software Analysis, Evolution and Reengineering (SANER'17). Piscataway, NJ: IEEE; 2017. p. 24–33.

29. rOpenSci Code of Conduct [Internet]. rOpenSci [cited 2019 Feb 14]. https://ropensci.org/code-of-conduct/.

30. SciPy Community. NumPy Code of Conduct [Internet]. SciPy Community [cited 2019 Feb 14]. https://www.numpy.org/devdocs/dev/conduct/code_of_conduct.html.

31. Code of Conduct [Internet]. Project Jupyter [cited 2019 Feb 14]. https://github.com/jupyter/governance/blob/master/conduct/code_of_conduct.md.

32. Ehmke, CA. Contributor Covenant [Internet]. [cited 2019 Feb 14]. https://www.contributor-covenant.org/.

33. SciPy Code of Conduct [Internet]. SciPy [cited 2019 Feb 14]. https://docs.scipy.org/doc/scipy/reference/dev/conduct/code_of_conduct.html.

34. Aurora V, Gardiner M. How to Respond to Code of Conduct Reports. Version 1.1 ed. Frame Shift Consulting LLC; 2019.

35. Steinmacher I, Conte T, Gerosa MA, Redmiles D. Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. CSCW 2015. New York City, NY: ACM Press; 2015. p. 1379–1382.

36. Ford D, Smith J, Guo PJ, Parnin C. Paradise Unplugged: Identifying Barriers for Female Participation on Stack Overflow. In: Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE'16). FSE 2016. New York City, NY: ACM Press; 2016. p. 846–857.

37. Baltes S, Dumani L, Treude C, Diehl S. The Evolution of Stack Overflow Posts: Reconstruction and Analysis. arXiv: 1811.00804 [Preprint]. 2018 [cited 2019 Mar 27]. https://arxiv.org/abs/1811.00804.

38. Ford D, Lustig K, Banks J, Parnin C. "We Don't Do That Here": How Collaborative Editing with Mentors Improves Engagement in Social Q&A Communities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI 2018. New York City, NY: ACM Press; 2018. Paper 608.

39. Steinmacher I, Wiese I, Conte T, Gerosa M, Redmiles D. The Hard Life of Open Source Software Project Newcomers. In: Proceedings of the 7th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE'14). New York City, NY: ACM Press; 2014. p. 72–78.

40. Steinmacher I, Treude C, Gerosa MA. Let Me In: Guidelines for the Successful Onboarding of Newcomers to Open Source Projects. IEEE Software. 2018; 36: 41–49. https://doi.org/10.1109/MS.2018.110162131

41. Doherty-Sneddon G, OMalley C, Garrod S, Anderson A, et al. Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. Journal of Experimental Psychology: Applied. 1997; 3(2):105–125. https://doi.org/10.1037/1076-898x.3.2.105

42. Gallupe RB, McKeen JD. Enhancing Computer-Mediated Communication: An experimental investigation into the use of a Group Decision Support System for face-to-face versus remote meetings. Information & Management. 1990; 18(1):1–13. https://doi.org/10.1016/0378-7206(90)90059-q

43. Nardi BA, Whittaker S. The place of face-to-face communication in distributed work. In: Hinds P, Kiesler S, editors. Distributed Work. Cambridge, MA: MIT Press; 2002. p. 83–110.

44. O'Malley C, Langton S, Anderson A, Doherty-Sneddon G, Bruce V. Comparison of face-to-face and video-mediated interaction. Interacting with Computers. 1996; 8(2):177–192. https://doi.org/10.1016/0953-5438(96)01027-2

45. Huppenkothen D, Arendt A, Hogg DW, Ram K, VanderPlas JT, Rokem A. Hack weeks as a model for data science education and collaboration. Proc National Academy of Sciences. 2018; 115(36):8872–8877. https://doi.org/10.1073/pnas.1717196115 PMID: 30127025

46. Cherubini M, Venolia G, DeLine R, Ko AJ. Let's Go to the Whiteboard: How and Why Software Developers Use Drawings. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07). New York City, NY: ACM Press; 2007. p. 557–566.

47. Fagerholm F, Guinea AS, Münch J, Borenstein J. The role of mentoring and project characteristics for onboarding in open source software projects. In: Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM'14). New York City, NY: ACM Press; 2014. Article 55.

48. Labuschagne A, Holmes R. Do Onboarding Programs Work? In: 2015 IEEE/ACM 12th Working Conference on Mining Software Repositories (MSR'15). Piscataway, NJ: IEEE; 2015. p. 381–385.

49. Barcomb A, Stol KJ, Riehle D, Fitzgerald B. Why Do Episodic Volunteers Stay in FLOSS Communities? In: Proceedings of the 41st International Conference on Software Engineering (ICSE'19). Piscataway, NJ: IEEE; 2019. p. 948–954.

50. Mäenpää H, Mäkinen S, Kilamo T, Mikkonen T, Männistö T, Ritala P. Organizing for openness: six models for developer involvement in hybrid OSS projects. Journal of Internet Services and Applications. 2018; 9(1): 17. https://doi.org/10.1186/s13174-018-0088-1

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules towards healthier research labs

**Fernando T. Maestre** *

Departamento de Biología y Geología, Física y Química Inorgánica, Universidad Rey Juan Carlos, Móstoles, Spain

* fernando.maestre@urjc.es

## Abstract

The negative effects of extremely competitive academic and research environments on the performance and health of researchers are well known and common worldwide. The prevalence of these effects, particularly among early career researchers, calls for a more humane and people-centered way of working within research labs. Although there is growing concern about the urgent need for a better life–work balance when doing science, there are not many examples about how this could be achieved in practice. In this article, I introduce 10 simple rules to make the working environment of research labs more nurturing, collaborative, and people-centered. These rules are directed towards existing and future principal investigators (PIs) but will be of interest to anyone working in a research lab and/or dealing with how to improve working conditions for scientists.



**Citation:** Maestre FT (2019) Ten simple rules towards healthier research labs. PLoS Comput Biol 15(4): e1006914. https://doi.org/10.1371/journal.pcbi.1006914

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** April 11, 2019



**Funding:** This work was supported by the European Research Council (ERC Grant Agreement 647038 [BIODESERT]). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The author has declared that no competing interests exist.

"We are all smart. Distinguish yourself by being kind."—Charles Gordon

## Introduction

Doing science often looks like a dream job, but many aspects of current scientific practice across the world make it a stressful activity. These include the shortage of scientific positions, the fear of being scooped by competing labs, the pressure to publish in high profile journals –-which are both a key indicator of success and prestige and crucial to secure positions/promotions–the uncertainty imposed by short-term contracts, and even the competition with other lab mates [1–5]. Scientists must also confront established practices that (i) force them to become workaholics if they want to get a permanent position and/or become successful [6], (ii) often promote harassment against women and minorities [7, 8], and (iii) emphasize success when failures (of experiments and/or simulations) and rejections (of articles, proposals, and job applications) are inseparable from the scientific endeavour [9, 10]. In addition, evaluation systems in place in many countries reward the number of publications in indexed journals over their quality [11–14], further pushing scientists to publish as much and as quickly as possible. This adds additional stress to researchers and promotes scientific malpractices such as nepotism and collusion, which negatively affect their well-being [11, 14–16]. With these antecedents, it is not surprising to read that many young and talented principal investigators (PIs) are frustrated [1] or to see surveys revealing that over 14% of non-PIs choose terms such as "competitive, stressful, or toxic" to describe their labs [17]. Even more worrying, recent studies and surveys have revealed that mental problems in academia are on the rise, with graduate students showing alarming rates of anxiety and depression [18, 19].

The issues described above call for a shift in scientific practice, which is urgently needed to protect scientists against health risks and abuses associated with extreme pressure and other negative habits that have plagued research labs for decades. Such a shift would also benefit the entire scientific enterprise by favoring more human and creative environments in which ideas can flourish and ground-breaking discoveries and innovations can be made, something that is at odds with the pressure to publish quickly and often [11, 15, 20]. Although there is growing concern about the urgent need for a better life–work balance and for creating healthier working habits [17–19], there are few examples of how this could be achieved in practice [e.g., 21–25]. Indeed, limited information and resources on this topic and lack of training on management and/or leadership skills by PIs are often invoked as barriers to effectively create healthier working environments in academic and research institutions [17, 25].

In this article, I present 10 simple rules based on my experience as a PI to make research labs nurturing, collaborative, and people-centered research environments. As leaders of research labs and role models for early career scientists, PIs, and particularly those that are well-established in their respective fields, have a critical role to play in promoting a shift towards creating healthier research environments. Therefore, these rules are directed toward existing and future PIs but will be of interest to anyone working in a research lab and/or dealing with how to improve working conditions for scientists.

## Rule 1: Promote the well-being of your lab members

We work more efficiently and are more creatively when we are happy. This is well known by psychologists studying productivity in the workplace across a wide range of jobs [e.g., 26–28]. The well-being of lab members must be a priority for PIs, who should devote important efforts to make research labs places where everyone can work in the best conditions possible while at the same time enjoying doing science. There is no unique path to achieve this but putting yourself in the situation of the others; being kind; banning all forms of harassment and discrimination within the lab; being sensitive when it comes to dealing with personal, family and health situations; and carefully listening to lab members regarding any matter related to their work can substantially improve the well-being of lab members. It is important to let your lab members know that you care about them and that you are here to listen to and to help them to overcome any issues that may negatively affect their work.

## Rule 2: Let people set their own schedules

As PIs, we should not strictly control lab members' schedules, and we should be flexible regarding their working preferences. Some people prefer to come early in the morning to have the afternoons free, whereas others prefer to do the opposite. Sometimes, it is more effective to stay at home when analyzing data and writing or to reconcile work and family obligations. PIs should facilitate these arrangements, because scientists should be evaluated by the outcome of their work rather than by the time they spend in their workplace (which for many researchers can be any spot with a computer and an internet connection). As Gandalf the Grey says, "All we have to decide is what to do with the time that is given us" [29], and it is the responsibility of lab members to use their working time wisely and to do whatever best works for them. Of course, when applying this rule, we must also keep in mind both differences among disciplines and the particular challenges faced by each lab. For example, graduate students and postdocs can work on an individual programming project or on the writing of a manuscript at home if they are more productive, but this may not work in collaborative projects requiring teammates to coordinate schedules.

From my experience, offering lab members this flexibility works very well for most people; it also helps graduate students and postdocs to learn how to manage their time effectively, something very important given the multiple tasks they will have to do when and if they become PIs. It must also be noted that offering lab members flexibility to set their own schedules does not remove our obligations as PIs to properly supervise them. We must hold periodic meetings with lab members to check the progress of their work. This is also very important to let them know that we care and are on top of what they do, as well as to discuss solutions when problems arise or when expectations are not met. Not doing so is indeed a source of frustration, particularly for graduate students [17].

## Rule 3: Gratitude is the sign of noble souls

Psychologists are well aware of the multiple benefits of being grateful [30, 31]. This not only has very positive knock-on effects on the work and personal well-being of lab members but also helps to build confidence and compromise among them. Showing our gratitude to lab members is important because their work, from the accounting done by administrative assistants to the data gathered by technicians or the writing of manuscripts by graduate students or postdocs, is crucial to ensure the smooth running of a research group. PIs can also show lab members how important their work is by providing rapid feedback to their requests, questions, and manuscript drafts. This is something that they really appreciate, particularly graduate students and postdocs, and contributes to boosting their motivation. Although at particularly busy periods it may not be possible to provide quick feedback to the request of lab members, trying to do so should always be our priority as PIs.

## Rule 4: Treat your lab members as your teammates

It is not uncommon to find labs with clearly established hierarchies and "top-down" approaches, particularly when it comes to the treatment of graduate students and technicians. Such an approach promotes toxic relationships and limits the capacity of lab members to think critically. As PIs, we must have the vision, set the research priorities for our labs, and have the last say on multiple matters. However, treating lab members as mere executors of our instructions rather than as colleagues that have an informed opinion about the work they do (and hence about how to improve it!) is a huge lost opportunity. We must listen to and take the opinion and advice of technicians, graduate students, and postdocs very seriously and often discuss with them ideas for projects and papers, lab procedures, and day-to-day issues affecting their work and well-being.

For this rule to work, PIs must also learn to delegate important work. Doing so relieves PIs of extra duties that other lab members can do more efficiently, such as doing chemical analyses on the lab or filling administrative forms. It also motivates lab members to become more engaged with lab projects and overall research objectives, thus contributing to teambuilding.

## Rule 5: Create a collaborative environment within your lab

Collaboration is a cornerstone of current scientific practice [32] that allows scientists to tackle ambitious, expensive, or multidisciplinary projects not amenable to a single lab. Doing science as a collective endeavour also brings multiple opportunities for learning and professional development, particularly for early career researchers. Therefore, as PIs, we must actively practice and foster collaborations within our labs, which also helps lab members to get along better with each other (something very important to maintain a happy and productive lab!). These collaborations also help to foster long-term relationships that can also be very fruitful for their professional development. Within-lab collaborations can be nurtured by setting up common lab projects, encouraging meetings and discussions involving all lab members, providing time and resources to develop side projects and/or ideas coming from them, conducting retreats and regular meetings outside the lab, and facilitating interactions between graduate students and postdocs. Establishing priorities and identifying needs in advance, knowing how to organize the work of everyone, and being gentle in the way we ask for help when needed also contribute to setting up effective collaborations within our labs.

Creating a collaborative, rather than competitive, environment within research labs not only helps everyone pull in the same direction but also fosters the motivation, productivity, and creativity of lab members. This also prepares them to set up collaborations with colleagues from other institutions, which are also very important for their career development.

## Rule 6: Remember that every lab member is unique

A key rule we must follow as PIs is not to compare our lab members to one another or with ourselves when we were students and/or postdocs. Comparing lab members will often result in increased stress and/or anxiety levels, reducing their performance and capabilities. Every person is different, and, as PIs, we should never forget that our major role as mentors is to foster everyone's capabilities and help them to fulfill their potential and professional ambitions. Therefore, we should make every effort to identify these goals and to support them by choosing appropriate projects and forging the right contacts. We must keep in mind that our objective as PIs is to help our lab members reach as far as they can and/or want, not as far as we want.

## Rule 7: Respect working hours, public holidays, and vacations

Working rules commonly in place in labs around the world often mean that academics work all day long, on weekends, and even during holidays [1, 6]. The stress associated with this excessive work without a life outside the lab is one of the main reasons behind the increase in mental problems in academia, particularly among early career researchers and young PIs [1, 19]. This also has many other deleterious effects on the health and well-being of researchers (see [33] for a review). Therefore, PIs should not expect lab members to work beyond normal hours, during weekends, and on holidays. We all face moments (e.g., deadlines for grant submissions, setup of large experiments, field campaigns) in which we must work hard. But this should be the exception, not the rule. Doing so is unsustainable in the long-term and contributes to generating expectations about the research environment that are neither realistic for many people nor desirable and/or healthy for the whole scientific community.

This rule can be seen as contradictory by junior PIs or those who are running labs that are short of labor and other resources, who are struggling with keeping a lab funded, or who are worried about tenure or establishing their reputation. We must also keep in mind the large differences that exist between different countries and cultures about what constitutes a "normal" working week, the length of annual holidays, and the pressures induced by the requirements to getting a job or being promoted. But even in these cases, it is important to remember that our working conditions are regulated by law and our contracts, and that working for long hours is not a sine qua non condition for being successful as a scientist (something that is intimately linked to our personal life [35]), as multiple examples from around the world illustrate [6, 24, 34].

Despite the importance of this rule for maintaining healthier research labs, as PIs we should also respect those lab members who choose to work for long hours because they feel that they must do so to be more productive, to secure a position in science, or because they have the ambition or the desire to be so. In the end, this is part of their freedom and autonomy (things are seen very differently from a permanent and/or well-established position) and we cannot forget that scientific productivity is important for the future career prospects of PhD students and postdocs. But at the same time, we should discourage these habits and advise them about the long-term ill effects that they may have on their health and well-being.

To gain the maximum benefit from this rule, PIs must also openly discuss and share with all lab members resources and experiences and/or tips to work more efficiently so they can maximize their productivity within normal working hours to avoid the need to work beyond them.

## Rule 8: Give credit where credit is due

We all have either experienced or heard about PIs who dictate authorship inclusion or order, or who insist on being authors on every paper produced by lab members, regardless of their contribution. This practice only benefits those in power, discourages effective collaborations, impedes the productivity and creativity of lab members, and fosters frustration and distrust among non-PIs. Therefore, it should be abolished. As PIs, we must openly discuss coauthorship issues with our lab members and train them on the importance of carefully evaluating the merits of coauthors before submitting publications. Failing to include meritorious coauthors or including undeserving coauthors can easily lead to frustrations and misunderstandings that must be avoided.

There are multiple ways we can give proper credit to PIs, including involving technicians in publications when they have contributed to them, leaving "senior" (e.g., last author) positions to postdocs when they had the idea of the study and are not first authors, declining authorship in articles in which we did not participate, and acknowledging in talks with colleagues, seminars, and scientific meetings the intellectual authorship of publications or ideas coming from our lab members.

## Rule 9: Destigmatize failure and celebrate success

Active scientists face rejection of their papers, grants, and job applications continuously, no matter what their career stage and status are [9, 10, 36]. Focusing on success while living under continuous rejection may put more pressure on the work of our graduate students and postdocs, increasing their frustration and anxiety levels when their articles or applications are rejected. And although rejection always hurts, scientists must embrace it as another (and important) part of their job [9, 10]. Initiatives to normalize rejection include the building of "a CV of failures" (see [37] for a great example), talking openly and sharing our experiences about rejection, and discussing with lab members the potential reasons for a particular rejection and how to avoid it the next time. Showing our lab members that rejection is the rule, rather than the exception, will help them to navigate the turbulent waters of research, reduce the prevalence of the "impostor syndrome" [38], and boost their self-confidence. And because successes are not so common, they must be properly celebrated when they happen. Fortunately, this is a usual practice in many labs that also contributes to the establishment of fruitful personal and professional interactions between lab members.

## Rule 10: Promote the professional development of your lab members

There is no single way this rule can be put into practice, because it may vary markedly among fields, countries, cultures, and personal situations. However, getting informed and openly discussing with lab members the pros and cons of all possible career options can help to do so. PIs should also allow time and resources (whenever available) to allow those lab members wishing to continue with a career in science to get trained in critical aspects of this job, such as experimental design, statistical analyses, and scientific writing. In addition, PIs should facilitate that graduate students and postdocs develop their own network of contacts, something that can be fostered by attending scientific meetings, by conducting research stays in other labs, and by participating in networks of scientists and specialist groups within scientific societies. Finally, PIs should also allow graduate students and postdocs to supervise BSc and MSc theses, respectively, on their own or under PI cosupervision, and offer postdocs the possibility of cosupervising new PhD students. By doing so, graduate students and postdocs acquire key experience on how to supervise the work of students, a critical task in academia; students get another view and critical inputs that end up improving their training and work; and PIs can have more time to do other important day-to-day issues that are needed to run a research lab. Furthermore, this action also effectively contributes to fostering collaborations and personal relationships between lab members.

## Concluding remarks

Strong competition promoted by the scarcity of funding and positions will continue to characterize science in many countries over the coming years. However, its negative effects on the performance and health of researchers [4, 5, 13, 17, 33] and the prevalence of such effects among new PIs, graduate students, and postdocs [1, 17–19] call for a more humane and people-centered way of working within research labs.

The 10 rules presented here can be adapted and refined to the circumstances of each research lab and can go a long way in improving working conditions and minimizing the negative effects of current scientific practice around the world. Embracing these rules could also help to alleviate other major problems faced by science today that can be exacerbated by extreme competition and the need to "be the first," such as publication of poor-quality data, reduced research standards, fraudulent behavior, and lack of reproducibility [11, 13–15, 39, 40]. For the rules presented here to work, as PIs, we should practice them on a daily basis. If we stay for long hours in the lab, work on the weekends, and spend our holidays doing fieldwork, attending conferences or catching up with the literature, how can we convince our lab members of the need to have a proper life–work balance?

The 10 rules described here may not work for everyone. There will always be scientists who prefer to sacrifice part of their personal lives or health to be as productive and successful as possible, even though working for long hours may not help them to do so. Some scientists work in countries where the institutional system and/or prevailing cultural practice make it very difficult to implement them. But even in these cases, the application of some of the rules presented (e.g., Rules 1–5 and 9) can alleviate many of the problems associated with unhealthy working habits. It is also important to remember that, in the end, our work is only part of our life, and that achieving a proper life–work balance will make us healthier, happier, and more productive in the long term [34, 35]. Reasons for this include a lower stress load, a higher capacity to concentrate on important tasks, and boosted energy, satisfaction, and motivation levels [33, 34].

As PIs, we have a major influence on how the people working with us will behave once they become PIs. Therefore, we must not only act with probity and apply sound scientific practices and decisions but do as much as we can to make our labs more humane, collaborative, and healthy. We must also use any influence we have beyond our own labs to change current scientific and institutional practices, something that potentially can benefit thousands of scientists worldwide.

To sum up, something that I always keep in mind as a marathon runner: a research career is like a long-distance race, in which running too fast in the first kilometers can lead to a drop-off of the race before reaching the finish line. Therefore, let's help our teammates to enjoy (and finish) the race by promoting a healthy and less stressful working environment.


## Acknowledgments

I would like to thank all the past and present members of my lab for shaping the rules presented here over the years and to my family for helping me to realize that there are more important things in life than work. I also would like to thank Miguel Berdugo, Matthew Bowker, Manuel Delgado-Baquerizo, David Eldridge, Pablo García-Palacios, Nicolas Gross, Rocío Hernández-Clemente, Mónica Ladrón de Guevara, Yoann Le Bagousse-Pinguet, Raúl Ochoa, César Plaza, José Luis Quero, Hugo Saiz, David Sánchez, Santiago Soliveres, and Enrique Valencia for their multiple comments and suggestions on previous drafts, which have substantially helped to improve this article. The idea of writing this article came after the responses obtained to a Twitter thread (https://twitter.com/ftmaestre/status/1053154881077415936), which I wrote after a very enjoyable conversation with Luis Rey.


## References


1. Powell K. Young, talented and fed-up: scientists tell their stories. Nature 2016; 538: 446–449. https://doi.org/10.1038/538446a PMID: 27786221
2. Woolston C. Paths less travelled. Nature 2018; 562: 611–614. https://doi.org/10.1038/d41586-018-07111-8 PMID: 30356203
3. Pearson H. Competition in biology: it's a scoop! Nature 2003; 426: 222–223. https://doi.org/10.1038/426222a PMID: 14628017
4. Gray Blix A, Cruise RJ, Mitchell BM, Blix GG. Occupational stress among university teachers. Educ Res 1994; 36: 157–169.
5. Kinman G, Jones F, Kinman R. The Well-being of the UK Academy, 1998–2004. Quality Higher Ed 2006; 12: 15–27.
6. Woolston C. Workplace habits: Full-time is full enough. Nature 2017; 546: 175–177.
7. Witze A. Sexual harassment is rife in the sciences, finds landmark US study. Nature 2018; 558: 352–353. https://doi.org/10.1038/d41586-018-05404-6 PMID: 29921850
8. Burke KA. Harassment in science. Am Sci 2017; 105: 262.
9. Topalidou I. Teach undergraduates that doing a PhD will require them to embrace failure. Nature 2018; https://doi.org/10.1038/d41586-018-06905-0
10. Parkes E. Scientific progress is built on failure. Nature 2019; https://doi.org/10.1038/d41586-019-00107-y
11. Huang F. Quality deficit belies the hype. Nature 2018; 564: S70–S71.
12. Hedding DW. Payouts push professors towards predatory journals. Nature 2019; 565: 267. https://doi.org/10.1038/d41586-019-00120-1 PMID: 30644448
13. Carson L, Bartneck C, Voges K. Over-competitiveness in Academia: A literature review. Disrup Sci Technol 2013; 1: 183–190.
14. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US states data. PLoS ONE 2010; 5: e10271. https://doi.org/10.1371/journal.pone.0010271 PMID: 20422014
15. Smaldino PE, McElreath R. The natural selection of bad science. R Soc Open Sci. 2016; 3:160384. https://doi.org/10.1098/rsos.160384 PMID: 27703703
16. Boulbes DR, Costello T, Baggerly K, Fan F, Wang R, Bhattacharya R, Ye X, Ellis LM. A survey on data reproducibility and the effect of publication process on the ethical reporting of laboratory research. Clin Cancer Res. 2018; 24:3447–3455. https://doi.org/10.1158/1078-0432.CCR-18-0227 PMID: 29643062
17. Van Noorden R. Some hard numbers on science's leadership problems. Nature 2018; 557: 294–296. https://doi.org/10.1038/d41586-018-05143-8 PMID: 29769686
18. Woolston C. Graduate survey: A love–hurt relationship. Nature 2017; 550: 549–552.

19. Evans TM, Bira L, Gastelum JB, Weiss LT, Vanderford NL. Evidence for a mental health crisis in graduate education. Nat Biotechnol 2018; 36:282–284. https://doi.org/10.1038/nbt.4089 PMID: 29509732
20. Foster JG, Rzhetsky A, Evans JA. Tradition and Innovation in Scientists' Research Strategies. Am Soc Rev 2015; 80: 875–908.
21. Norris D, Dirnagl U, Zigmond MJ, Thompson-Peer K, Chow TT. Health tips for research groups. Nature 2018; 557: 302–304. https://doi.org/10.1038/d41586-018-05146-5 PMID: 29769687
22. Kwok R. How lab heads can learn to lead. Nature 2018; 557: 457–459. https://doi.org/10.1038/d41586-018-05156-3 PMID: 29769688
23. Woolston C. Feeling overwhelmed by academia? You are not alone. Nature 2018; 557: 129–131.
24. Maestre FT. Seven steps towards health and happiness in the lab. Nature 2018; https://doi.org/10.1038/d41586-018-07514-7
25. Antes A. First law of leadership: be human first, scientist second. Nature 2018; 563: 601. https://doi.org/10.1038/d41586-018-07530-7 PMID: 30479391
26. Oswald AJ, Proto E, Sgroi D. Happiness and productivity. J Labor Econ 2015; 33: 789–822.
27. Fitzgerald CJ, Danner KM. Evolution in the office: how evolutionary psychology can increase employee health, happiness, and productivity. Evol Psychol 2012; 10: 770–781. PMID: 23253786
28. Walsh LC, Boehm JK, Lyubomirsky S. Does happiness promote career success? Revisiting the evidence. J Career Assess 2018; 26: 199–219.
29. Tolkien JRR. The Lord of the Rings. The Fellowship of the Ring ( London: George Allen & Unwin, 1954).
30. Kaplan S, Bradley-Geist JC, Ahmad A, Anderson A, Hargrove A, Lindsay A. A test of two positive psychology interventions to increase employee well-being. J Bus Psychol 2014; 29: 367–380.
31. Fagley NS, Adler MG. Appreciation: a spiritual path to finding value and meaning in the workplace. J Manage Spirit Rel 2012; 9: 167–187.
32. Vermeulen N, Parker JN, Penders B. Understanding life together: a brief history of collaboration in biology. Endeavour 2013; 37: 162–71. https://doi.org/10.1016/j.endeavour.2013.03.001 PMID: 23578694
33. Clark MA, Michel JS, Zhdanova L, Pui SY, Baltes BB. All work and no play? A meta-analytic examination of the correlates and outcomes of workaholism. J Manage 2016; 42: 1836–1873.
34. Rosen J. How a hobby can boost researchers' productivity and creativity. Nature 2018; 558: 475–477. https://doi.org/10.1038/d41586-018-05449-7 PMID: 29915421
35. Clark A, Sousa B. How to Be a Happy Academic: A Guide to Being Effective in Research, Writing and Teaching ( London: SAGE Publications, 2018).
36. Cassey P, Blackburn TM. Publication and rejection among successful ecologists. BioScience 2004; 54: 234–239.
37. Hrala J. This Princeton professor's CV of failures is something we should all learn from. Science Alert, 2018, November 5. Available from: https://www.sciencealert.com/why-creating-a-cv-of-failures-is-good-Princeton-professor-viral.
38. Woolston C. Faking it. Nature 2016; 529: 555–557. PMID: 26824109
39. Schulz JB, Cookson MR, Hausmann L. The impact of fraudulent and irreproducible data to the translational research crisis–solutions and implementation. J Neurochem 2016; 139: 253–270. https://doi.org/10.1111/jnc.13844 PMID: 27797406
40. Fanelli D. Opinion: Is science really facing a reproducibility crisis, and do we need it to?. Proc Natl Acad Sci USA 2018; 115: 2628–2631. https://doi.org/10.1073/pnas.1708272114 PMID: 29531051

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for avoiding and resolving conflicts with your colleagues

**Fran Lewitter**[1¶]*, **Philip E. Bourne**[2], **Teresa K. Attwood**[3]

**1** Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, United States of America, **2** Data Science Institute & Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America, **3** School of Computer Science, The University of Manchester, Manchester, United Kingdom

¶ This author is retired.
* lewitter@iscb.org





**Citation:** Lewitter F, Bourne PE, Attwood TK (2019) Ten Simple Rules for avoiding and resolving conflicts with your colleagues. PLoS Comput Biol 15(1): e1006708. https://doi.org/10.1371/journal.pcbi.1006708

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** January 18, 2019



**Funding:** The authors received no funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

During the course of our personal and professional lives, we spend a significant amount of time communicating with others. In fact, communication is one of the most important, but possibly also one of the hardest, things we do, having the power to bring individuals and communities together or create divisions. Getting it right is therefore crucial.

Modern technologies have had a significant impact on the ways in which we are now able to communicate, allowing us to share our thoughts with colleagues, family, or friends at the click of a button. But communicating more quickly does not always result in better communication—the technologies we use often divorce us from the visual clues that are so crucial to understanding each other's true meaning and make it easy to misinterpret each other's real intentions. For this reason, our interactions can sometimes be unexpectedly difficult or can go unaccountably wrong.

Given that communication is vital to the health and productivity of relationships, how can we best make our interactions work, and how can we resolve situations when they arise? The following are 10 simple rules based on our experience that we hope will help. Many of these rules can apply to the kinds of communication we may have with colleagues, family, or friends. However, we focus this article on the professional environment: we begin with suggestions to help avoid disagreements or to help stop them turning into serious conflicts; we then reflect on steps that might help to resolve situations that have become confrontational.

Most interactions with colleagues are cordial and are working towards a common goal. Sometimes, however, because of differing views, misinterpretation of something said, or just because you're having a bad day, communications can go awry and become heated; from this point, without resolution, awkward situations can quickly escalate. Practicing effective communication skills before a confrontation arises, or during a confrontation, is the topic of this article. For more general ideas about engaging in successful collaborations, see [1]. To delve further into the area of conflict management in the work environment, see [2, 3]. To keep this contribution manageable, we have confined ourselves to peer-to-peer communication and not considered a larger ecosystem of interactions in which conflict occurs. We feel that is a separate contribution that should be written.

## Rule 1: Always treat people with equality and respect

Whether interacting with your peers or not, treat people courteously. Don't prejudge individuals based on their rank or perceived academic abilities, or worse, their gender, race, or sexual orientation. Be polite, and treat everyone equally and fairly.

## Rule 2: Seriously consider and respect others' views

You may not always agree with a colleague's views, but remember, your colleagues may have different and valuable perspectives based on their past experience; moreover, based on their cultural background, they may have different sensitivities and differences in communication style that go far beyond language differences. Don't be dismissive of their ideas or values, and contemplate any cultural aspects that should be considered. Remember that the primary reason for interacting with your colleague was to achieve a common goal. Always keep that in mind, and try to work on ways to successfully reach that goal together.

## Rule 3: If you disagree with someone, say so and explain why

During the course of a discussion, you may find that you disagree—perhaps quite strongly—with a colleague. If this happens, remain calm and professional; don't let anger or resentment build up. In short, communicate calmly and early to prevent confrontation. It's important to address points of friction as they arise. Be straightforward about what concerns you or what you object to. Give reasons for your point of view—state the facts, not your opinions. In trying to defuse a situation, or to prevent an impasse, it may be helpful to include additional people in the conversation; to be constructive, you should also try to propose a solution to your disagreement.

## Rule 4: Make sure you are on the same page

It is particularly easy to misinterpret communications when discussions are conducted via email or another asynchronous medium. If, because you're busy, your communications are short and abrupt, they may just come across as rude; on the other hand, if your messages are long and emphatic, they may seem rather dogmatic. To reduce the chance of misinterpretation and to ensure that what you mean to convey isn't lost, it's important to be clear in your written communication and to choose your words carefully—be particularly aware of the vocabulary and grammar you use. This is especially important if the other person is not a native speaker of the language you're using and may miss nuances in what you are saying. During a confrontation, backing up any written communication with a verbal follow up, however hard, is advisable.

## Rule 5: Pause before you press "Send"

Whether you are writing or speaking to a colleague, think before you do so. If you're upset and plan to send an email, it is advisable to write a draft and put it aside for a while. Sleep on it! Take time to ensure that the content of the message is really what you want to convey and that it is measured in tone and objective (as mentioned in Rule 4, email communication can very easily be misinterpreted). The same goes for verbal communication. If you are irritated or frustrated, take time to compose yourself and to formulate rational arguments before speaking to your colleague. In our experience, this is the most likely way to prevent an escalation in the situation.

## Rule 6: Apologize when you do something wrong

We all make mistakes—a sharp word in a meeting, an email sent in haste, a spontaneous tweet. If you have had a disagreement about an issue or treated someone disrespectfully, there's nothing wrong (but everything right) in offering a sincere apology, preferably in writing. Whether you want to continue working with the person you wronged or not, it's best to admit you erred. This is more likely to earn you respect from your colleagues, and in the long run, it will be respect that goes a long way to define you as a scientist [4].

## Rule 7: Engage in an honest nonconfrontational dialogue

If a disagreement has escalated into a conflict, distance yourself from your emotions and document the conflict: note what was said or done to you, when and where, and how it made you feel. Having a written record is extremely useful. Ask to have a conversation with your colleague (whether face-to-face, by telephone, Skype, etc.). Agree on the issue to be discussed; it may be helpful to provide a concise summary to avoid misconceptions. Present your position clearly and listen carefully to the response. Ask for clarification if you don't understand what is said to you. Don't let emotions enter into the discussion; avoid raising your voice. If you feel confident, calm, and safe enough to do so, address the behavior. Don't give your colleague permission to bully you. Let him or her know that the behavior is offensive or, at the very least, making you uncomfortable. If possible, ask where the behavior has come from; try to get to the real issue and help your colleague resolve his or her anger.

## Rule 8: Know who to turn to if you need impartial advice

If you find yourself being confronted by a colleague who uses inappropriate language or who makes you feel threatened, don't reply in kind. Empower yourself by finding out your organization's policy on bullying and harassment (you may be protected by their policies or code of conduct) and making yourself aware of the law in your country regarding bullying and harassment in the workplace (in many countries, bullying or offensive, intimidating, and/or abusive behavior of colleagues toward you may actually be unlawful). Take action by reporting the behavior to someone outside the situation—make the impact of the behavior, in terms of how it made you feel, very clear. If you and your colleague are from the same department, the department head may be able to offer advice to help resolve the conflict. However, this option needs to be considered in light of the individuals involved. If not approached carefully, it may backfire and cause further resentment from your colleague, escalating rather than defusing the situation. In this case, you may need to go to a higher level individual outside your department. Many institutions and organizations have ombudspersons to handle conflicts that get out of hand. Find out what your institution or organization offers. Alternatively, you may have an impartial mentor or colleague to whom you may turn for advice, but if a professional skilled in dealing with conflict is available, that is preferred.

## Rule 9: Know when it's worth fighting for your point of view

There's an old saying: "Don't beat your head against the wall. If you do, it feels much better when you stop." If you are in an argument that is not moving towards resolution, walk away. In the best of circumstances, it is desirable that you and your colleague can simply agree to disagree and move on to other issues, still maintaining mutual respect.

## Rule 10: If the problem is unresolvable, distance yourself

Unfortunately, not all conflicts can be resolved to everyone's liking. It may be necessary at some point to distance yourself from the source of the problem. That may include resigning from a position, leaving a research project, or shifting your focus to other activities—a sad situation to be in but one that is preferable to continuing to endure an unpleasant environment.

Like it or not, you will encounter a spectrum of conflicts throughout your scientific career. Those range from disagreements as part of the daily scientific discourse—those are what drive science forward—to the unpleasant, in which personalities and associated emotions clash in disagreeable ways. How you handle the conflict spectrum will go a long way to defining you as a scientist. It's a small and intertwined community. What goes around comes around. Act appropriately at all times with, we hope, these rules as your guide.

## References


1. Bourne PE, Barbour V (2011) Ten simple rules for building and maintaining a scientific reputation. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1002108 PMID: 21738465
2. Gallo A (2017) HBR Guide to Dealing with Conflict. Boston: Harvard Business Review Press.
3. Proksch S (2016) Conflict Management. Switzerland: Springer International Publishing.
4. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.0030044 PMID: 17397252

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for international short-term research stays

**Diego A. Forero[1,2]*, Sandra Lopez-Leon[3], George P. Patrinos[4,5]**

**1** Laboratory of NeuroPsychiatric Genetics, Biomedical Sciences Research Group, School of Medicine, Universidad Antonio Nariño, Bogotá, Colombia, **2** PhD Program in Health Sciences, School of Medicine, Universidad Antonio Nariño, Bogotá, Colombia, **3** Global Drug Development, Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, United States of America, **4** Department of Pharmacy, University of Patras School of Health Sciences, Patras, Greece, **5** Department of Pathology, College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, United Arab Emirates

* diego.forero@uan.edu.co



OPEN ACCESS

**Citation:** Forero DA, Lopez-Leon S, Patrinos GP (2017) Ten simple rules for international short-term research stays. PLoS Comput Biol 13(12): e1005832. https://doi.org/10.1371/journal.pcbi.1005832

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** December 7, 2017



**Funding:** DAF has been supported by research grants from VCTI-UAN and Colciencias. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Because science is a global endeavor, international mobility is common among researchers and academics around the world [1,2]. Short-term research stays (from a few weeks to a few months), which do not involve a change of employer or affiliation, are the main form of international mobility [3]. This type of "brain circulation" [4] increases collaborations, creates networks, improves career prospects, facilitates the generation of high-impact publications, gives access to international funding, and nourishes ideas through exposure to different methods and scientific skills [5–7]. In this article, we present 10 simple rules for better and more productive experiences of international scientific mobility, which might be helpful for scientists in all stages of their careers, particularly for MSc and PhD students [8].

## Rule 1: Select your host institution

The affinity in academic and research topics between the host and the guest laboratories is a priority [9,10]. Other affinities (e.g., nationality, language, or culture) come second. However, do take into account that countries with little mobility embrace diversity less easily [11]. Studies have shown that the main drivers of choosing a host institution are having an outstanding faculty or research team; the infrastructure and facilities; and the expertise, excellence, prestige, or high quality of the foreign institution in a certain area [11–13]. There are several ways of finding an adequate host. Simple ways are to send emails to authors who publish work in your field of interest, to ask supervisors and your network, or to attend short courses and conferences. In addition, you can search websites of universities, academic centers, regulatory agencies, and industry. Some scientific societies have a section in which possibilities for short stays are announced, and an industry posts their intern research positions in traditional job posting websites (e.g., indeed.com, linkedin.com, monster.com, experteer.com). Contacting the international office of the host university would help with some administrative processes and for exploring additional options. In many cases, the administrative staff of the host department is quite helpful in supporting logistical aspects of the international visit.

## Rule 2: Plan ahead carefully

Planning for a research stay abroad takes time in order to organize in advance the different aspects; namely, adequate time in the host institution, plane tickets, accommodation, visas, insurance and permissions, details of the experiments to be carried out, and rotations in other laboratories and/or centers, among others [9,14]. It is imperative that the necessary paperwork for the exchange student and/or staff be executed well in advance. Scientists should not leave their country before these documents are signed by both parties. All details (time, remuneration in certain cases, casualty and legal liability insurance, among others) should be laid out in the contract. This requires adequate coordination between the members of the host and guest laboratories and institutions. It is key to have a clear understanding of the timelines of the main regulatory processes required in the host institution, such as approval by the local research ethics committee.

Once you identify the host country, it is essential that you visit the website of the embassy and/or consulate in your location and that you contact the international office of your institution to find out the requirements and the time it takes to issue the required travel documents. In addition, the embassy should be able to provide extra information, such as sources for funding, and advise you if any (extra) vaccines are needed. Make sure your health insurance covers you in the host country; otherwise, get international insurance. If you are planning on driving, and your license is not valid in the host country, obtain an international driving license (usually valid for up to one year).

## Rule 3: Define funding needs and sources for your research stay in advance

Obtaining funding for mobility is the biggest challenge for researchers [3]. Costs of reagents or the use of specialized equipment requires funding to cover those expenses, which might be charged to the host or the guest's laboratory. Sometimes, this is covered by overheads available in host institutions as part of their research grants. In several cases, the results from a research stay become pilot data for future collaborative projects aimed at obtaining more funding [9]. If neither your institution nor the host institution provides funding, you can search for grants given by other institutions, include the international stay in your grant proposals, or consider saving and paying for your stay yourself. A survey of almost 9,000 PhD students found that more than 20% used personal savings to finance stays abroad [15].

## Rule 4: Respect the organization of the host institution

There are differences between institutions in the way that departments are organized, and some of these differences might be amplified because of dissimilarities between countries (and more importantly, cultures). Respect for the organization of the host institution might facilitate an adequate development of the activities to be carried out by the guest scientist. Knowing details of the organization, structure, and dynamics of the host institution helps in the process of integrating the guest researcher. Before you travel, you can talk to several people working in the department and to past visiting researchers to understand what to expect.

During your stay, follow the wisdom of the saying "When in Rome, do as the Romans do." Do not take things for granted, and when in doubt, ask. Never share data from the organization with outside people and do not use materials and/or programs or analyze data without prior written authorization.

## Rule 5: Be prepared for integration into the host laboratory

It is key that you look for a potential direct mentor or adviser (in several cases, it is a scientist different from the principal investigator [PI] of the host laboratory) who takes a key role in your constant supervision and academic support from an early moment. You are encouraged to attend laboratory meetings in advance of your stay—e.g., using videoconferencing tools—to have a better understanding of the dynamics of the host laboratory. You need to have a good understanding of the language of the country of the host laboratory to facilitate adequate communication with the other members. In some cases, English is the main option, but in several countries, other languages are needed. Integration into the host laboratory would facilitate the completion of key goals defined in advance of the research stay and an adequate reporting of findings and advances (e.g., presentations in laboratory and department meetings and preparation of manuscripts, among others).

Expanding your scientific knowledge is the priority. You should not see the international experience as merely a means to boost your resume or as a paid leisure trip. Leisure should be seen as secondary, and extra time should be set aside for it either by leaving some hours at the end of the workday or by adding extra days to your stay.

A key point for a successful research stay is that the scientific activities to be carried out at the host institution remain the main priority [10,11,14]. This means that there is a need (particularly for students) for an adequate articulation to the guest's home laboratory of the activities to be carried out at the host institution. Take into account that part of the collaborative work can be finalized in your own country, so focus on the activities for which one needs to be physically present face-to-face—such as scientific discussions and the use of databases, computer programs, laboratory samples, materials, and instruments that can only be used at the host institution.

## Rule 6: Define authorship and acknowledgments in advance for the products of your research stay

As with other type of collaborative endeavors, it is fundamental to define in advance the authorship of the possible publications and/or conferences resulting from the research stay. If authorship and conference presenters are not clarified beforehand, this usually leads to unpleasant situations between research groups and may lead to discontinuation of collaborations [9]. Take into account that guidelines for assigning authorship vary between fields; therefore, make sure to understand what to expect. Find out what the requirements are to be considered an author. Define who is going to be first author, how the order of the author list will be determined, and who is going to be the corresponding author. In some areas, being the last author means that you were the PI, while in others it means that you contributed the least. Although listing authors in order of involvement seems straightforward, a study showed that more than two-thirds of 919 corresponding authors disagreed with their coauthor order [16].

In addition to defining the authorship of publications and/or conference proceedings in advance, other types of acknowledgment of the collaborations are important [11], such as presentations in scientific events and in other types of documents (MSc and PhD thesis, final reports of projects, future grant applications, among others) [17].

## Rule 7: Learn from the differences

In several cases, one of the most valuable experiences from international mobility is learning different styles of doing science [18]. It is also a good opportunity to practice other languages and to have an immersion in other cultures. Going to lectures or laboratory meetings and journal clubs is a great way to learn from the academic environment of the host institution.

Go with the mentality that you are the one who has to adapt to their culture and way of working; do not expect people to adapt to you. A good resource to understand the differences between your culture and the culture you are visiting is the book *The Culture Map* [19] and the series of interviews that *Science* published for international scientists (http://www.sciencemag.org/careers/2011/08/international-mobility). If the language and culture are completely different, then there may be benefits of having a cultural coach who can explain unfamiliar hierarchies, conflict management, and body language.

## Rule 8: Try to resolve problems in an adequate way during your research stay

Inconveniences between research groups and/or members are frequent, and these conflicts might be of particular relevance in research stays of international guests [10,11]. An early identification of possible inconveniences is key in order to manage them in a transparent and cordial manner [9]. When dealing with other cultures, communication is essential because many of the problems arise because of misunderstandings. If something is bothering you, or if you expect something, say it. Let them know that you are open to their feedback and that you would like to know if there is something that was expected from you. Keep in mind that things that are obvious in your country may not be so in your host country; therefore, there is a need for you to be flexible and open-minded and to try to adapt as much as possible.

## Rule 9: Explore other options for carrying out collaborations

The use of videoconferencing, file sharing services [20], and webinars facilitates some facets of international collaborations because these resources are less expensive and complex than international flights [21]. These strategies, which might save costs and personnel time and effort, might be of particular importance for researchers from resource-limited countries [22]. Consider collaborations with other sectors, including industry, contract research organizations, regulatory agencies, and other universities. Diverse teams that are able to complement each other often lead to the development of innovation and creativity [23].

## Rule 10: You can be a host as well

Invite a researcher to your institution. Having an international researcher in your institution can bring the same advantages to the host and to the visitor, including collaborations, networks, and exposure to different methods and scientific skills; plus, visiting guests could give lectures and bring experience and new ideas into your laboratory or institution. Involve your supervisors, legal, and human resources in the idea from an early stage because they will be essential in helping with the paperwork and accommodations. When you write a research grant, you can consider including an international visitor.

## Conclusion

Please note that every single rule and all advice given in this article can also be applied by the host researcher. Several of the rules are related to general aspects of scientific collaborations, but they have particular characteristics inherent to the dynamics of short-term international scientific mobility. The advancement of international communication and geographical mobility has made "brain circulation" accessible. However, when visiting research institutions abroad, there are many details that need to be taken into account that are easily missed. Therefore, these rules will be an essential source of guidance when planning your visit. Congratulations on your next international short stay; you will have one of the best experiences in your academic life.

## Acknowledgments

The authors acknowledge the insights generated from a number of experiences in different countries and institutions and thank the help from Albert Ruiter in the revision of the manuscript.

## References


1. Wagner CS, Park HW, Leydesdorff L (2015) The Continuing Growth of Global Cooperation Networks in Research: A Conundrum for National Governments. PLoS ONE 10: e0131816. https://doi.org/10.1371/journal.pone.0131816 PMID: 26196296
2. Ioannidis JP (2004) Global estimates of high-level brain drain and deficit. FASEB J 18: 936–939. https://doi.org/10.1096/fj.03-1394lfe PMID: 15173104
3. Franzoni C, Scellato G, Stephan P (2012) Foreign-born scientists: mobility patterns for 16 countries. Nat Biotechnol 30: 1250–1253. https://doi.org/10.1038/nbt.2449 PMID: 23222798
4. Petersen AM, Puliga M (2017) High-skilled labour mobility in Europe before and after the 2004 enlargement. J R Soc Interface 14.
5. Van Noorden R (2012) Global mobility: Science on the move. Nature 490: 326–329. https://doi.org/10.1038/490326a PMID: 23075963
6. Cañibano C, Otamendi J, Andújar I (2008) Measuring and assessing researcher mobility from CV analysis: the case of the Ramón y Cajal programme in Spain. Research Evaluation 17: 17–31.
7. Scellato G, Franzoni C, Stephan P (2015) Migrant scientists and international networks. Research Policy 44: 108–120.
8. Forero DA, Moore JH (2016) Considerations for higher efficiency and productivity in research activities. BioData Min 9: 35. https://doi.org/10.1186/s13040-016-0115-3 PMID: 27833658
9. de Grijs R (2015) Ten Simple Rules for Establishing International Research Collaborations. PLoS Comput Biol 11: e1004311. https://doi.org/10.1371/journal.pcbi.1004311 PMID: 26447799
10. Gu J, Bourne PE (2007) Ten simple rules for graduate students. PLoS Comput Biol 3: e229. https://doi.org/10.1371/journal.pcbi.0030229 PMID: 18052537
11. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3: e44. https://doi.org/10.1371/journal.pcbi.0030044 PMID: 17397252
12. Appelt S, van Beuzekom B, Galindo-Rueda F, de Pinho R (2015) Which factors influence the international mobility of research scientists? Global Mobility of Research Scientists: The Economics of Who Goes Where and Why: 177–214.
13. Franzoni C, Scellato G, Stephan P (2015) International mobility of research scientists: Lessons from GlobSci. Global mobility of research scientists, the economics of who goes where and why, Amsterdam: Elsevier: 35–65.
14. Yu M, Kuo YM (2017) Ten simple rules to make the most out of your undergraduate research career. PLoS Comput Biol 13: e1005484. https://doi.org/10.1371/journal.pcbi.1005484 PMID: 28472033
15. Schiermeier Q (2011) Career choices: The mobility imperative. Nature 470: 563–564. PMID: 21381250
16. Ilakovac V, Fister K, Marusic M, Marusic A (2007) Reliability of disclosure forms of authors' contributions. CMAJ 176: 41–46. https://doi.org/10.1503/cmaj.060687 PMID: 17200389
17. Jolly M, Fletcher AC, Bourne PE (2012) Ten simple rules to protect your intellectual property. PLoS Comput Biol 8: e1002766. https://doi.org/10.1371/journal.pcbi.1002766 PMID: 23144604
18. Knapp B, Bardenet R, Bernabeu MO, Bordas R, Bruna M, et al. (2015) Ten simple rules for a successful cross-disciplinary collaboration. PLoS Comput Biol 11: e1004214. https://doi.org/10.1371/journal.pcbi.1004214 PMID: 25928184
19. Meyer E (2014) The culture map: Breaking through the invisible boundaries of global business: PublicAffairs.
20. Boland MR, Karczewski KJ, Tatonetti NP (2017) Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing. PLoS Comput Biol 13: e1005278. https://doi.org/10.1371/journal.pcbi.1005278 PMID: 28103227
21. Gichora NN, Fatumo SA, Ngara MV, Chelbat N, Ramdayal K, et al. (2010) Ten simple rules for organizing a virtual conference—anywhere. PLoS Comput Biol 6: e1000650. https://doi.org/10.1371/journal.pcbi.1000650 PMID: 20195548
22. Moreno E, Gutierrez JM (2008) Ten simple rules for aspiring scientists in a low-income country. PLoS Comput Biol 4: e1000024. https://doi.org/10.1371/journal.pcbi.1000024 PMID: 18437198
23. Patrinos GP, Katsila T (2016) Pharmacogenomics education and research at the Department of Pharmacy, University of Patras, Greece. Pharmacogenomics.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing

**Mary Regina Boland[1,2,3,4]*, Konrad J. Karczewski[5,6], Nicholas P. Tatonetti[1,2,3,4]**

**1** Department of Biomedical Informatics, Columbia University, New York, New York, United States of America, **2** Department of Systems Biology, Columbia University, New York, New York, United States of America, **3** Department of Medicine, Columbia University, New York, New York, United States of America, **4** Observational Health Data Sciences and Informatics, Columbia University, New York, New York, United States of America, **5** Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **6** Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America

* mary.boland@columbia.edu

Open access, open data, and software are critical for advancing science and enabling collaboration across multiple institutions and throughout the world. Despite near universal recognition of its importance, major barriers still exist to sharing raw data, software, and research products throughout the scientific community. Many of these barriers vary by specialty [1], increasing the difficulties for interdisciplinary and/or translational researchers to engage in collaborative research. Multi-site collaborations are vital for increasing both the impact and the generalizability of research results. However, they often present unique data sharing challenges. We discuss enabling multi-site collaborations through enhanced data sharing in this set of Ten Simple Rules.

Collaboration is an essential component of research [2] that takes many forms, including internal (across departments within a single institution) and external collaborations (across institutions). However, multi-site collaborations with more than two institutions encounter more complex challenges because of institutional-specific restrictions and guidelines [3]. Vicens and Bourne focus on collaborators working together on a shared research grant [4]. They do not discuss the specific complexities of multi-site collaborations and the vital need for enhanced data sharing in the multi-site and large-scale collaboration context, in which participants may or may not have the same funding source and/or research grant.

While challenging, multi-site collaborations are equally rewarding and result in increased research productivity [5, 6]. One highly successful multi-site and translational collaboration is the Electronic Medical Records and Genomics (eMERGE) network (URL: https://emerge.mc.vanderbilt.edu/) initiated in 2007 [7]. The eMERGE network links biorepository data with clinical information from Electronic Health Records (EHRs). They were able to find novel associations and replicate many known associations between genetic variants and clinical phenotypes that would have been more difficult without the collaboration [8]. eMERGE members also collaborated with other consortiums and networks, including the Alzheimer's Disease Genetics Consortium [9] and the NINDS Stroke Genetics Network [10], to name a few. Other successful collaborations include OHDSI: Observational Health Data Sciences and Informatics (http://www.ohdsi.org/), which builds off of the methodology from the Observational Medical Outcomes Partnership (OMOP) [11], and CIRCLE: Clinical Informatics Research Collaborative (http://circleinformatics.org/). In genetics, there are many consortiums, including ExAC: The Exome Aggregation Consortium (http://exac.broadinstitute.org/), the 1000 Genomes Project Consortium (http://www.1000genomes.org/), the Australian BioGRID (https://www.biogrid.org.au/), The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/),

OPEN ACCESS

**Citation:** Boland MR, Karczewski KJ, Tatonetti NP (2017) Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing. PLoS Comput Biol 13(1): e1005278. doi:10.1371/journal.pcbi.1005278

**Published:** January 19, 2017



**Funding:** MRB was supported by NLM T15 LM00707 from Jul 2014–Jun 2016 and by the NCATS, NIH, through TL1 TR000082, formerly the NCRR, TL1 RR024158 from Jul 2016–Jun 2017. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

Genotype-Tissue Expression Portal (GTEx: http://www.gtexportal.org/home/), and Encyclopedia of DNA Elements at UCSC (ENCODE: https://genome.ucsc.edu/ENCODE/) among others.

Based on our experiences as both users and participants in collaborations, we present ten simple rules on how to enable multi-site collaborations within the scientific community through enhanced data sharing. The rules focus on understanding privacy constraints, utilizing proper platforms to facilitate data sharing, thinking in global terms, and encouraging researcher engagement through incentives. We present these ten rules in the form of a pictograph of modern life (**Fig 1**), and we provide a table of example sources and sites that can be referred to for each of the ten rules (**Table 1**). Please note that this table is not meant to be exhaustive, only to provide some sample resources of use to the research community.

## Definitions

In this paper, we use the term "research product" to include all results from research. This includes algorithms, developed software tools, databases, raw source data, cleaned data, and various metadata generated as a result of the research activity. We differentiate this from "data," which comprises the primary "facts and statistics collected together for analysis" for that particular collaboration. Therefore, data could include genetic data or clinical data. By these definitions, developed software tools are not "data" but "research products." Novel genetic sequences collected for analysis would be considered "raw source data," which is a type of "research product."

## Rule 1: Make Software Open-Source

The cornerstone of facilitating multi-site collaborations is to enhance data sharing and make software open-source [12]. By allowing the source code to be open, researchers allow others to both reproduce their work and build upon it in novel ways. To engage in multi-site collaborations, it is necessary for collaborators to have access to code in a repository that is shared among collaborators (although, this could be private and not open to the general public). When the study is complete and the paper is under review and/or published, a stable copy of the code should be made available to the general public. Internal sharing allows the code to be developed, while public sharing of a stable version allows the code to be refined and built upon by others.

Many researchers still limit access to their work despite the known advantages of making software open-source upon publication (e.g., higher impact publications [5]). For example, they allow users to interact with their algorithm by inputting data and receiving results on a web platform, while the backend algorithm often remains inaccessible. Masum et al. advocate the reuse of existing code in their Ten Simple Rules for cultivating open science [13]. However, this is often easier said than done. As long as the backend algorithms remain hidden, open science will not be possible. Therefore, it is essential for researchers interested in participating in multi-site collaborations to make their software code and algorithms open. Because making software truly "open" can be complex, Prlic and Proctor provide Ten Simple Rules to assist researchers in making their software open-source [12]. Truly open-source software is an essential component in collaborations [13]. Openness also has advantages for the researchers themselves. With more eyes on the source code, others within the community can refine the code, leading to greater identification and correction of errors. There are several methods for sharing software code. If you use the R platform, then libraries can be shared with the entire open-source community via CRAN (https://cran.r-project.org/) and bioconductor, which is specifically for biologically related algorithms (https://www.bioconductor.org/). Code can also be shared on Github with issue trackers for error detection.

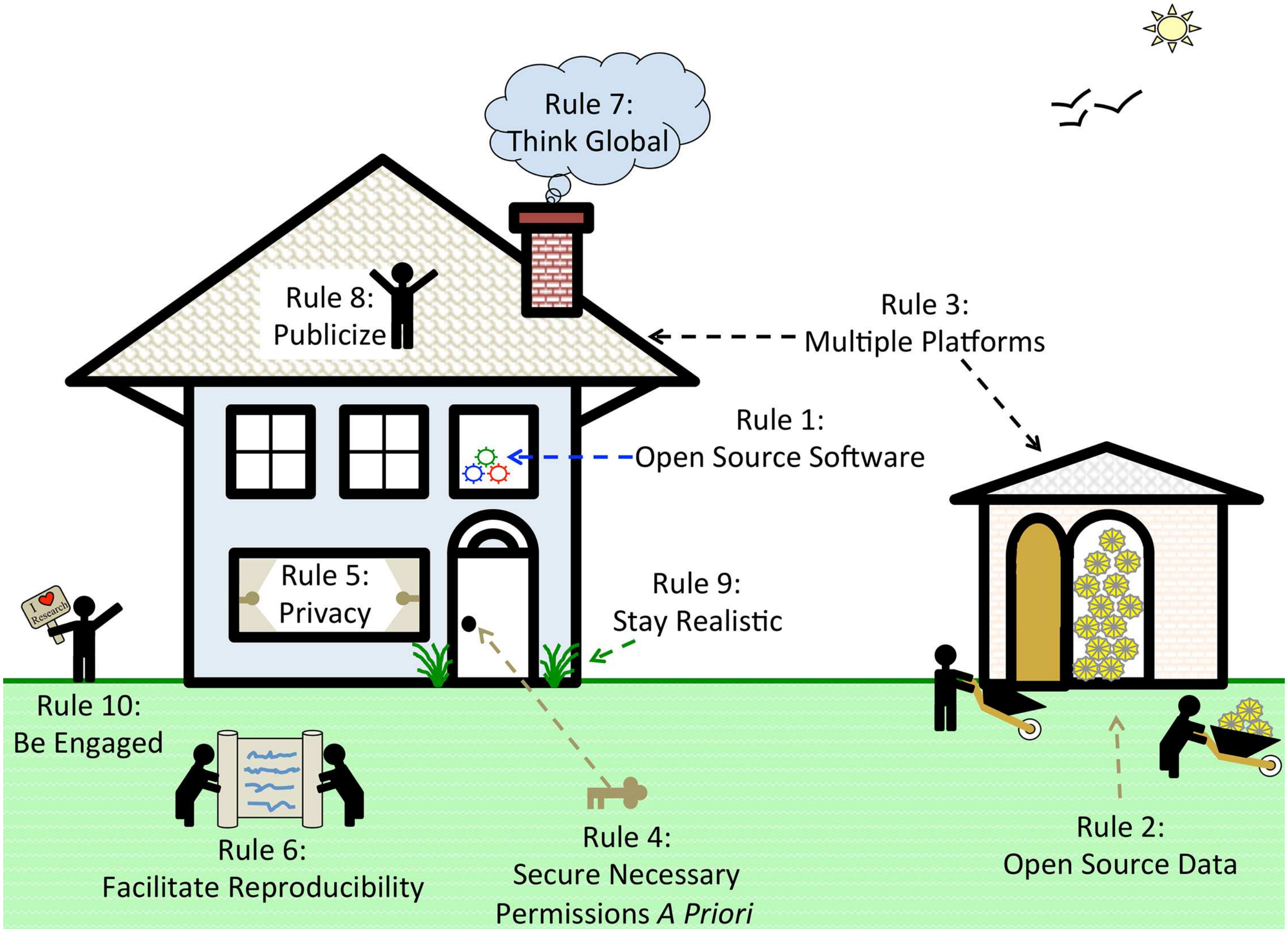



**Fig 1. Modern life context for the ten simple rules.** This figure provides a framework for understanding how the "Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing" can be translated into easily understood modern life concepts. **Rule 1** is Open-Source Software. The openness is signified by a window to a room filled with algorithms that are represented by gears. **Rule 2** involves making the source data available whenever possible. Source data can be very useful for researchers. However, data are often housed in institutions and are not publicly accessible. These files are often stored externally; therefore, we depict this as a shed or storehouse of data, which, if possible, should be provided to research collaborators. **Rule 3** is to "use multiple platforms to share research products." This increases the chances that other researchers will find and be able to utilize your research product—this is represented by multiple locations (i.e., shed and house). **Rule 4** involves the need to secure all necessary permissions a priori. Many datasets have data use agreements that restrict usage. These restrictions can sometimes prevent researchers from performing certain types of analyses or publishing in certain journals (e.g., journals that require all data to be openly accessible); therefore, we represent this rule as a key that can lock or unlock the door of your research. **Rule 5** discusses the privacy issues that surround source data. Researchers need to understand what they can and cannot do (i.e., the privacy rules) with their data. Privacy often requires allowing certain users to have access to sections of data while restricting access to other sections of data. Researchers need to understand what can and cannot be revealed about their data (i.e., when to open and close the curtains). **Rule 6** is to facilitate reproducibility whenever possible. Since communication is the forte of reproducibility, we depicted it as two researchers sharing a giant scroll, because data documentation is required and is often substantial. **Rule 7** is to "think global." We conceptualize this as a cloud. This cloud allows the research property (i.e., the house and shed) to be accessed across large distances. **Rule 8** is to publicize your work. Think of it as "shouting from the rooftops." Publicizing is critical for enabling other researchers to access your research product. **Rule 9** is to "stay realistic." It is important for researchers to "stay grounded" and resist the urge to overstate the claims made by their research. **Rule 10** is to be engaged, and this is depicted as a person waving an "I heart research" sign. It is vitally important to stay engaged and enthusiastic about one's research. This enables you to draw others to care about your research.

doi:10.1371/journal.pcbi.1005278.g001

**Table 1. Example sources and sites for each of the ten simple rules.**

| Rule | Example | Site |
|---|---|---|
| **Rule 1: Make Software Open-Source** | | |
| | Github | https://github.com |
| | CRAN | https://cran.r-project.org |
| | Bioconductor | https://www.bioconductor.org |
| **Rule 2: Provide Open-Source Data** (When Possible) | | |
| *Deposit Source Data in Appropriate Repositories* | | |
| | Sequence Read Archive (SRA) | https://www.ncbi.nlm.nih.gov/sra |
| | Gene Expression Omnibus (GEO) | https://www.ncbi.nlm.nih.gov/geo |
| | ClinVar | https://www.ncbi.nlm.nih.gov/clinvar |
| *Consider Middle-Ground Data Sharing Approaches for Sensitive Data* | | |
| | dbGaP | https://www.ncbi.nlm.nih.gov/gap |
| | Shared Health Research Information Network (SHRINE) | https://catalyst.harvard.edu/services/shrine |
| | BioGrid Australia | https://www.biogrid.org.au |
| **Rule 3: Use Multiple Platforms to Share Research Products** | | |
| | Figshare | https://figshare.com |
| | Github | https://github.com |
| | ExAC Browser | http://exac.broadinstitute.org |
| | Google Forums | |
| **Rule 4: Secure Necessary Permissions/Data Use Agreements A Priori** | | |
| *Guides for Creating a DUA* | | |
| | Department of Health and Human Services Best Practice Guide for DUA | http://www.hhs.gov/ocio/eplc/EPLC%20Archive%20Documents/55-Data%20Use%20Agreement%20(DUA)/eplc_dua_practices_guide.pdf |
| | Health Care Systems Research Network DUA Toolkit | http://www.hcsrn.org/en/Tools%20&%20Materials/GrantsContracting/HCSRN_DUAToolkit.pdf |
| *Example DUAs* | | |
| | NASA DUA | http://above.nasa.gov/Documents/NGA_Data_Access_Agreement_new.pdf |
| | SEER-MEDICARE DUA | https://healthcaredelivery.cancer.gov/seermedicare/obtain/seerdua.docx |
| **Rule 5: Know the Privacy Rules for Your Data** | | |
| | Health Insurance Portability and Accountability Act (HIPAA) | http://www.hhs.gov/hipaa/for-professionals/privacy |
| **Rule 6: Facilitate Reproducibility** | | |
| *Resources for Increasing Research Reproducibility* | | |
| | MetaSub Research Integrity and Reproducibility | http://metasub.org/research-integrity-and-reproducibility/ |
| | Reproducibility and Open Science Working Group—GitHub | http://uwescience.github.io/reproducible/guidelines.html https://github.com/uwescience/reproducible |
| *Example Projects with Assessed Reproducibility* | | |
| | eMERGE PheKB | https://phekb.org/network-associations/emerge |
| **Rule 7: Think Global** | | |
| *Guides for Collaborating Globally* | | |
| | National Academies "Collaborating with Foreign Partners to Meet Global Challenges"Resources | http://sites.nationalacademies.org/PGA/PGA_041691 |
| | Global Alliance for Genomics and Health | http://genomicsandhealth.org/work-products-demonstration-projects/catalogue-global-activities-international-genomic-data-initiati |
| | The Global Strategy of the US Department of Health and Human Services | http://www.hhs.gov/sites/default/files/hhs-global-strategy.pdf |
| *Examples of Successful International Projects* | | |
| | Human Fertility Database | http://www.humanfertility.org/cgi-bin/main.php |
| | Human Mortality Database | http://www.mortality.org |
| **Rule 8: Publicize Your Work** | | |
| *Research Without Novelty Requirement* | | |

(*Continued*)

Table 1. (*Continued*)

| Rule | Example | Site |
|---|---|---|
| | *PLOS ONE* | http://journals.plos.org/plosone |
| | *Scientific Reports* | http://www.nature.com/srep |
| | *Cell Reports* | http://www.cell.com/cell-reports/home |
| *Data Resources (Web Browsers, Databases)* | | |
| | *Scientific Data* | http://www.nature.com/sdata |
| | *Database* | https://database.oxfordjournals.org |
| *Pure Open Science Research (all data must be open)* | | |
| | F1000 | https://f1000research.com |
| **Rule 9: Stay Realistic** | | |
| | Retraction Watch | retractionwatch.com |
| **Rule 10: Be Engaged** | | |
| *Resources to Facilitate Researcher Engagement* | | |
| | KNAER Creating Partnerships: Learning New Ways to Connect | http://www.knaer-recrae.ca/blog-news-events |
| *Example Projects with Researcher Engagement* | | |
| | STAN | http://mc-stan.org |
| | STAN "swag" | http://mc-stan.org/shop |

doi:10.1371/journal.pcbi.1005278.t001

## Rule 2: Provide Open-Source Data

### Deposit Source Data in Appropriate Repositories

Whenever possible, it is important to make source data available. Openness benefits your collaborators by allowing them to perform additional analyses easily. Source data could include not only processed or cleaned data used in algorithms but also raw data files. These files can often be very large; therefore, they are often stored in some external site or data warehouse. The National Center for Biotechnology Information (NCBI) maintains the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra) and the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/); both are great places to deposit source data, if appropriate.

In addition to raw data files, it is also helpful to provide intermediate data files at various stages of processing. If comparing your results to those in the literature, it can also be useful to provide a meta-analysis with publications (along with PubMed IDs) that detail those publications that support and refute the results you obtained.

Data sharing is vitally important for multi-site collaborations by allowing researchers to compare results from across vastly different study populations, which increases the generalizability of the findings [14]. While a multi-site research project is still ongoing, data can be shared in a private shared space until all necessary data quality checks have been conducted and the findings have been published. After publication, data can be deposited in GEO, SRA, ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and any other domain-specific sites that are appropriate for source data deposition.

### Consider Middle-Ground Data Sharing Approaches for Sensitive Data

Raw source data is not always fully shareable with the public. This can be because of data use restrictions (see rule 4) or privacy concerns (see rule 5). Alternative mechanisms exist for sharing portions of data with the research community. For example, the database for Genotypes and Phenotypes or dbGaP (https://www.ncbi.nlm.nih.gov/gap) provides data holders with two levels of access: open and controlled. The open selection allows for broad release of nonsensitive data online, whereas the controlled release allows sensitive datasets to be shared with other investigators, provided certain restrictions are met. This increases the ability for researchers to share portions of their data that would not be shareable otherwise.

In addition to the restricted data sharing option provided by dbGaP, others have looked at ways of developing middle-ground approaches for sharing sensitive raw data or metadata. Several of these mid-level approaches use Federated Access systems that allow researchers to query databases containing sensitive data while preventing direct access to the data itself. An example within the United States is the Shared Health Research Information Network (SHRINE), which provides a Federated system that is HIPAA compliant [15]. International groups have also seen success in this area. BioGrid Australia (https://www.biogrid.org.au/) allows researchers to access hundreds of thousands of health records through a linked data platform where individual data holders maintain control of their data [16]. Researchers can then be provided with authorized access to certain elements within the data while restricting access to private sections of the medical data. These mid-level approaches facilitate collaboration both within the institution (i.e., across departments) and across institutions by allowing researchers to access sensitive data indirectly. They can even match patients to similar patients (for association analyses) while maintaining stringent privacy constraints [17]. Others provide summary statistics computed over large cohorts (e.g., ExAC browser/database), which maintains privacy while providing others with important information about the populations that can be used in subsequent analyses and comparisons.

## Rule 3: Use Multiple Platforms to Share Research Products

To collaborate with researchers from different backgrounds, it is often necessary to use multiple platforms when sharing data (as different disciplines often have different policies). Using multiple platforms allows individuals from diverse backgrounds to have access to your research product. General phrases like "open data" and "open science" are phrases used commonly in the research community but provide little direction [13]. Research products take many different forms, including 1) raw source data regardless of collection type (e.g., health data, genomic data, survey data, and epidemiological data), 2) software code (mentioned in rule 1), and 3) metadata elements and results of computations used to generate figures published in scientific research. Some data types cannot be fully shared (e.g., EHR data—see rule 5), but most algorithms and summary results/statistics are shareable.

Each of these types of open data necessitates a different platform for data sharing. Figshare (https://figshare.com/) allows users to share data involving published figures. Github (https://github.com/) allows users to share code that is in development or published. For code that is well developed, open-source packages can be created, for example, an R library, which can be deposited in CRAN or bioconductor. R libraries can be shared immediately on github without any code checking—this is advisable for code that is still in development. However, when code is finalized, it can be submitted to bioconductor as an R library. Approved libraries are vetted to ensure that the code works well. Vignettes are also good to write to help new users get used to the R package. When collaborating across multiple sites, it is also important to have vignettes and sample source data to help users learn how to use the code even if R is not your language of choice. Data formats, differences among formats, and programming languages are important to consider when sharing data across multiple platforms. Different platforms often have different required formats. While it may seem tedious to translate code, source data, and documentation across multiple formats and data schemas, it can be very helpful, and it will increase the number of users that will find your data and results interesting.

To facilitate communication among members of a collaborative effort, there are many options, including Google forums and wiki webpages, among others. Others have specially designed websites for the sole purpose of allowing users to browse and download the data directly; one such website is the ExAC Browser (http://exac.broadinstitute.org/), which integrates data obtained from 17 different consortiums (http://exac.broadinstitute.org/about) [18].

## Rule 4: Secure Necessary Permissions/Data Use Agreements A Priori

Some datasets have provisos that affect publication, and these need to be addressed a priori. For example, the ability for researchers to publish an algorithm that uses a Government dataset can depend on the department that generated the data. For example, certain National Aeronautics and Space Administration (NASA) datasets stipulate that data usage requires users to add certain NASA employees to subsequent publications. This is an important stipulation. Others may disallow the deposition of data into an "open" platform as part of their data use agreements (http://above.nasa.gov/Documents/NGA_Data_Access_Agreement_new.pdf). These stipulations can hinder researchers attempting to produce transparent science.

Other datasets have data use agreements as an added layer to ensure that patients are protected. For example, the Surveillance, Epidemiology, and End Results (SEER) dataset linked with Medicare (i.e., SEER-Medicare dataset) requires that users submit the intended publication to their offices for pre-submission approval. This can seem burdensome to researchers; however it is a condition of the data use agreement and, therefore, must be complied with. Researchers need to be aware of all provisos when including such data in their studies. Before publishing, or providing data in any type of platform whether open, restricted, or closed, it is important to secure all necessary provisions and data use agreements.

## Rule 5: Know the Privacy Rules for Your Data

Data come with many caveats. For this reason, it is important to understand what you can and cannot do (i.e., the privacy rules) with your data. Keeping and maintaining data privacy is different from data use agreements (DUA, see rule 4). For example, data that is not sensitive may have restrictive DUAs for other reasons (e.g., data from a collaborator in industry). Also, privacy rules often involve your own source data, whereas DUAs become necessary when using data from collaborators or a government source.

Certain datasets, e.g., genomic and EHR data, may be impossible to fully publish on an open platform due to the Health Insurance Portability and Accountability Act (HIPAA) privacy rules and other privacy concerns related to patient re-identifiability (http://www.hhs.gov/hipaa/for-professionals/privacy/). Therefore, it is important to know the privacy stipulations of **all** data used in your collaborations and how this affects the ability to share results among members of the team (especially when members of the team are at different institutions). Methods that anonymize patient information while allowing patient-level data sharing may be the way of the future [19]. However, institutional-specific policies and/or country-specific laws can limit or prevent usage of such methods. This is an important item to consider and discuss with all collaborators at the outset of any collaboration. We discuss some methods that can be used to provide some forms of sensitive data in a shareable federated space in rule 2.

## Rule 6: Facilitate Reproducibility

Another aspect of both data sharing and enabling multi-site collaborations is reproducibility. Sandve et al. provide Ten Simple Rules for facilitating research reproducibility in general [20]. Keeping track of research results and how data were generated is vital for reproducibility [20].

This site-level record keeping becomes vital when engaging in multi-site collaborations. If one aspect of a methodology is not conducted in the same way at one site, the overall results can be affected in drastic ways. In other words, reproducibility is a core requirement for successful collaborations.

In genetics and computational biology, the issue of standardizing results from across different types of gene sequencing platforms is a major issue [21]. Researchers that use a mixture of clinical and genetic data (for Phenome-Wide Association Studies, PheWAS [22]) often depend on local EHR terminology systems for identifying patient populations. Therefore, standard phenotype definitions are required and must be harmonized across multiple sites to ensure that the definitions are accurate at each site [23]. Several multi-site collaborations have developed platforms that provide links to all necessary documentation, code, and data schemas to help facilitate this process [24], including the eMERGE network. This step is integral to data sharing and enabling multi-site collaborations.

## Rule 7: Think Global

The importance of thinking globally cannot be overstated. Health care, genetics, climate, and all aspects of science affect the world as a whole. Therefore, it is important to think globally when performing scientific research. Most software languages are designed to be agnostic to the local language of the country. However, understanding and using these languages requires adequate documentation and user manuals to be provided in the local languages of the programmers/implementers. Despite this, open-source languages often provide user manuals in certain languages. For example, R is a popular open-source language yet has official documented translations in only four languages: English, Russian, German, and Chinese (https://www.r-project.org/other-docs.html). Problems can surface when collaborators in different regions run into difficulties with running R. This affects data sharing on a global scale and should be considered when collaborating on an international venue.

Translational mechanisms may also be necessary to understand and to harmonize country-specific terminology. This is especially important as definitions for obesity and many psychiatric conditions vary widely across the globe [25]. Even seemingly simple biological features (e.g., tall versus short) can be difficult to translate in global terms. For example, an average height Norwegian may appear to be tall in a different country. Translating biological features to common absolute metrics (e.g., height) helps to alleviate ambiguities that can occur from categorical variables. Certain diseases, especially psychiatric conditions, are extremely important to study at the multi-site level to increase the generalizability of the results [14]. However, psychiatric conditions are more difficult to translate without a thorough knowledge of how the condition is defined in the underlying country or region [25]. Solutions often involve using concrete measures, e.g., brain imaging analysis, versus subjective measures such as depression presence or absence [14].

There are many layers to thinking on a global scale. There are mechanical differences (i.e., the software language and documentation) and also the conceptual differences (i.e., country- or region-specific medical definitions). Organizations such as the World Health Organization work tirelessly to integrate different conceptual interpretations of diseases into a standard guideline. Using these guidelines and not a country-specific guideline helps your research work reach the broader scientific community.

Several groups have successfully integrated data across multiple countries and provided their data in an open form. The Max Planck Institute for Demographic Research (MPIDR) in Germany collaborated with two separate groups to produce two databases containing international data. Both datasets contain integrated results from over 30 countries. Additionally, all finished data (after cleaning) is made available to users in an open format via two specially designed databases: the Human Fertility Database (http://www.humanfertility.org/cgi-bin/main.php) [26] and the Human Mortality Database (http://www.mortality.org/) [27]. Only cleaned data are returned to users in a standardized format, allowing users to easily compare countries with one another. The MPIDR collaborated with the Vienna Institute of Demography (Austria) in creating the Human Fertility Database and the University of California, Berkeley for the Human Mortality Database. They provide a good example of a group that successfully harmonized definitions across countries by overcoming international barriers, and they provided data back to researchers in an easily useable and standardized format. The group provides detailed descriptions of how they harmonized various timescales across countries in a methods document (http://www.humanfertility.org/Docs/methods.pdf) that could easily be submitted as a research report (see Rule 6).

## Rule 8: Publicize Your Work

Publishing all aspects of your work in the appropriate venues is vital for maintaining a multi-site collaboration. This enables each aspect of your research to be assessed by appropriate peer reviewers. Publishing different aspects of your work in separate papers in separate journals allows your contributions to be seen by those most able to learn from your work. Remember, it is important to make your research work available to those who can benefit from your results. Depending on your findings, this can include methodologists, clinicians, epidemiologists, geneticists, and others.

New journals have been developed recently to facilitate open science, which are focused on certain aspects of research. For instance, there are several journals that do not require novelty as a requirement such as *PLOS ONE*, *Scientific Reports*, and *Cell Reports*. These journals are good choices for research results that may be part of a larger research project or collaborative but are not inherently novel. Other journals, such as *Scientific Data* and *Database*, are good choices for publishing a resource containing your collected research source data. It is often advisable to publish in data-focused journals simultaneously with an algorithm or results-focused paper that highlights the novel aspects of your research. In some cases, data can be published afterwards if it is part of a large collaborative and the database or user-interface is in production at the time that the main contribution is published.

Publishing in multiple venues is highly important for those engaged in multi-site collaborations, because these projects often involve a tremendous investment of time and resources from across many different organizations. Therefore, it is vital to highlight each and every research contribution that the collaboration has generated to facilitate further engagement from the community. If you are able to provide all raw source data on an open platform, there are new journals designed specifically to facilitate open science such as F1000 (https://f1000research.com/) that may be worth considering. F1000 is also a great source for intermediate results such as posters, which collaborators may have presented at various conferences while working towards the final finished paper. After publication, some collaborative groups effectively utilize blogging (both macro and micro) to communicate with other researchers and the general public. However, it is also important not to overstate the claims in any paper submission/publication or media regarding that publication but to stay focused on the individual contribution of that particular work.

## Rule 9: Stay Realistic, but Aim High

When performing quality research, and collaborating with others, it is important not to overstate the claims of your research—either in publication or online. It is vitally important to resist the urge to overstate the claims and to remain both humble and grounded. This is critical in collaborations because if a researcher overstates the claim in a paper, or worse, shares data publicly that he or she is unable to do legally (e.g., via the stipulations in a DUA), then the paper may be retracted. This could result in irreparable damage to the collaborative group.

This rule also links back to rule 2—making the source data available. This allows others in the research community to check your work interactively, which can help prevent overstating research claims [28]. A site exists that posts retracted journal articles on a public forum, retractionwatch.com. The site includes not only instances of plagiarism and fabrication of data but also papers that are retracted due to human error on the part of an experiment (e.g., a protocol was not followed exactly as specified in the paper) or on the part of the analysis (e.g., the wrong type of statistical test was performed, making the conclusions not substantiated by the data).

So, stay realistic, but do not be afraid to challenge the status quo. Some of the most respected research today was research that challenged the current understanding of the leading scientists at that point in time; this includes the seminal works on Pangaea and even that DNA is composed of a double helix. These concepts were earth-shattering at the time and could have been completely wrong, but the researchers backing them were not afraid to make their theories, data, and results public. These are the things that change science. So, remain humble, do not intentionally overstate the claims of your research, but at the same time do not be afraid to challenge the current mindset and way of thinking. You may be completely off, or you may just be a groundbreaking innovator.

## Rule 10: Be Engaged

Be engaged with those using your research, your data, and your code. Communicate with them using various software social platforms—Github, figshare, and so forth. Respond readily when users have questions and concerns. Attempt to follow the motto—release early, release often. Engage with researchers in non-traditional ways. For example, several collaborative efforts have created their own gear, e.g., t-shirts, to engage the community. One such collaborative is the open-source statistical modeling language—STAN (http://mc-stan.org/). They have created their own line of STAN “swag” (http://mc-stan.org/shop/) to facilitate user engagement. Communicate often with the research community to convince them your research is worth caring about. The bottom line in collaboration is to care deeply about your research. If you care and you make it known that you care deeply about the problem, then it becomes possible to convince others that your research is important.

## Concluding Remarks

Collaborations, especially large, multi-site collaborations, contain a lot of pitfalls that must be overcome. In this paper, we present ten simple rules that will help researchers share their data and methods to facilitate successful and meaningful multi-site collaborations. We describe these rules and highlight several successful multi-site collaborations.

## References

1. Reichman OJ, Jones MB, Schildhauer MP. Challenges and opportunities of open data in ecology. Science. 2011; 331(6018):703–5. doi: 10.1126/science.1197962 PMID: 21311007
2. Bozeman B, Fay D, Slade CP. Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. The Journal of Technology Transfer. 2013; 38(1):1–67.

3. Brown P, Morello-Frosch R, Brody JG, Altman RG, Rudel RA, Senier L, et al., editors. IRB challenges in multi-partner community-based participatory research. the American Sociological Association annual meeting; Sheraton Boston the Boston Marriott Copley Place; 2008.
4. Vicens Q, Bourne PE. Ten simple rules for a successful collaboration. PLoS Comput Biol. 2007; 3(3): e44. doi: 10.1371/journal.pcbi.0030044 PMID: 17397252
5. Jones BF, Wuchty S, Uzzi B. Multi-university research teams: Shifting impact, geography, and stratification in science. science. 2008; 322(5905):1259–62. doi: 10.1126/science.1158357 PMID: 18845711
6. Börner K, Contractor N, Falk-Krzesinski HJ, Fiore SM, Hall KL, Keyton J, et al. A multi-level systems perspective for the science of team science. Science Translational Medicine. 2010; 2(49):49cm24–49cm24. doi: 10.1126/scitranslmed.3001399 PMID: 20844283
7. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The electronic medical records and genomics (eMERGE) network: past, present, and future. Genetics in Medicine. 2013; 15 (10):761–71. doi: 10.1038/gim.2013.72 PMID: 23743551
8. Feng Q, Wei W, Chung C, Levinson R, Bastarache L, Denny J, et al. The effect of genetic variation in PCSK9 on the LDL-cholesterol response to statin therapy. The pharmacogenomics journal. 2016.
9. Karch CM, Ezerskiy LA, Bertelsen S, Goate AM. Alzheimer's Disease Risk Polymorphisms Regulate Gene Expression in the ZCWPW1 and the CELF1 Loci. PloS ONE. 2016; 11(2):e0148717. doi: 10.1371/journal.pone.0148717 PMID: 26919393
10. Malik R, Traylor M, Pulit SL, Bevan S, Hopewell JC, Holliday EG, et al. Low-frequency and common genetic variation in ischemic stroke The METASTROKE collaboration. Neurology. 2016; 86(13):1217–26. doi: 10.1212/WNL.0000000000002528 PMID: 26935894
11. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Annals of internal medicine. 2010; 153(9):600–6. doi: 10.7326/0003-4819-153-9-201011020-00010 PMID: 21041580
12. Prlić A, Procter JB. Ten Simple Rules for the Open Development of Scientific Software. PLoS Comput Biol. 2012; 8(12):e1002802. doi: 10.1371/journal.pcbi.1002802 PMID: 23236269
13. Masum H, Rao A, Good BM, Todd MH, Edwards AM, Chan L, et al. Ten Simple Rules for Cultivating Open Science and Collaborative R&D. PLoS Comput Biol. 2013; 9(9):e1003244. doi: 10.1371/journal.pcbi.1003244 PMID: 24086123
14. Pearlson G. Multisite Collaborations and Large Databases in Psychiatric Neuroimaging: Advantages, Problems, and Challenges. Schizophrenia Bulletin. 2009; 35(1):1–2. doi: 10.1093/schbul/sbn166 PMID: 19023121
15. Weber GM, Murphy SN, McMurry AJ, MacFadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories. Journal of the American Medical Informatics Association. 2009; 16(5):624–30. doi: 10.1197/jamia.M3191 PMID: 19567788
16. Merriel RB, Gibbs P, O'Brien TJ, Hibbert M. BioGrid Australia facilitates collaborative medical and bioinformatics research across hospitals and medical research institutes by linking data from diverse disease and data types. Human mutation. 2011; 32(5):517–25. doi: 10.1002/humu.21437 PMID: 21309032
17. Boyle DIR, Rafael N, editors. BioGrid Australia and GRHANITE: Privacy-Protecing Subject Matching. HIC; 2011.
18. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536(7616):285–91. http://www.nature.com/nature/journal/v536/n7616/abs/nature19057.html—supplementary-information. doi: 10.1038/nature19057 PMID: 27535533
19. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. bmj. 2015; 350: h1139. doi: 10.1136/bmj.h1139 PMID: 25794882
20. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol. 2013; 9(10):e1003285. doi: 10.1371/journal.pcbi.1003285 PMID: 24204232
21. Weis B. Standardizing global gene expression analysis between laboratories and across platforms. Nature methods. 2005; 2(5):351–6. doi: 10.1038/nmeth754 PMID: 15846362
22. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. Bioinformatics. 2010; 26(9):1205–10. doi: 10.1093/bioinformatics/btq126 PMID: 20335276
23. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network.

Journal of the American Medical Informatics Association. 2013; 20(e1):e147–e54. doi: 10.1136/amiajnl-2012-000896 PMID: 23531748

24. Pathak J, Wang J, Kashyap S, Basford M, Li R, Masys DR, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. Journal of the American Medical Informatics Association. 2011; 18(4):376–86. doi: 10.1136/amiajnl-2010-000061 PMID: 21597104

25. Diener E, Oishi S, Lucas RE. Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. Annual review of psychology. 2003; 54(1):403–25.

26. Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). 2016; http://www.humanfertility.org (data accessed on October 6, 2016).

27. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). 2016; http://www.mortality.org or http://www.humanmortality.de (data accessed on October 6, 2016).

28. Senn SJ. Overstating the evidence–double counting in meta-analysis and related problems. BMC Medical Research Methodology. 2009; 9(1):1.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Establishing International Research Collaborations

**Richard de Grijs**[1,2]*

**1** Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing, China, **2** East Asian Regional Office of Astronomy for Development (EA-ROAD), International Astronomical Union

* grijs@pku.edu.cn



OPEN ACCESS

**Citation:** de Grijs R (2015) Ten Simple Rules for Establishing International Research Collaborations. PLoS Comput Biol 11(10): e1004311. doi:10.1371/journal.pcbi.1004311

**Editor:** Scott Markel, Accelrys, UNITED STATES

**Published:** October 8, 2015



**Funding:** The author received no specific funding for this article. The author acknowledges support through grants from both major national funders and nonprofits, currently grant No. 11373010 from the National Natural Science Foundation of China.

**Competing Interests:** The author has declared that no competing interests exist.

Successful modern research collaborations increasingly include scientists based in different countries. This is partially driven by the need to engage with interdisciplinary science, access innovative approaches to problem solving, and acquire expertise beyond that which your own research group covers. It is also a great way to establish a worldwide network of colleagues with a variety of backgrounds—scientific, cultural, or otherwise. While international collaborations can be very rewarding, both professionally and from a personal perspective, they come with distinct difficulties and pitfalls that one should be aware of a priori. Nevertheless, cultivating an acute awareness of these issues will likely offer rich returns to internationally minded scientists, given that international research collaborations continue to expand, and many are now being established beyond the traditional power players, the North American and European research communities [1].

## Rule 1: Clarify Why You Might Want to Start an International Research Collaboration

Much has been written about the need for interdisciplinarity and cross-fertilization of ideas and methodologies in contemporary science [2]. Although many of the motivations as to why one might want to reach out to colleagues for help and assistance are equally valid for domestic as for international collaborations, pursuing this goal in an international context requires one to consider aspects that may not be as important in a domestic context. You may wonder why there is a need for international collaborations, particularly in fields in which the geographical location of the research project does not play an important role. Indeed, many research communities are split along national boundaries, if not formally then often in practice, as dictated by the strategic goals of the main national funding agencies. Foreign research groups may therefore offer access to complementary perspectives and insights, experience, and skills. Different national research priorities may also have given rise to country-specific differences in terms of the availability of resources or equipment beyond what you can access locally, regionally, or even domestically. International collaborations are indeed essential to tackling the grand challenges of our time [1].

The need to establish international collaborations may, in particular, apply in the context of sharing students, given the enormous differences in student numbers among the different national research communities. It is usually considered professionally advantageous to spend some time abroad during the early stages of a research career (when funding for such exchanges is fairly easily obtained), thus making exchanges of students and junior scientists a particularly interesting vehicle by which to establish international research collaborations. And since current technological advances facilitate easy communication, geographically distributed teams are not necessarily at a disadvantage.

In addition, many international collaborations are established following chance meetings between like-minded individuals, often at scientific meetings. Sharing your research progress with international collaborators with whom you have built up a personal relationship can indeed be very rewarding. In addition to pursuing quality research, the non-tangible benefits may include the development of a widening perspective and learning about different cultural aspects. You may make friends for life, who you will continue to encounter at meetings throughout your career.

## Rule 2: Consider the Characteristics Your International Collaborator Must Have

In addition to considering your motivation to establish an international research collaboration, you should also clarify the desired characteristics of your potential collaborator. This is particularly important if cultural differences or a language barrier may make communication more challenging. In this context, it is interesting to note that even today, many large international networks are split along common-language lines [1]. Consider asking local, trusted colleagues whether they could provide insights as regards your potential collaborator's reputation of reliability. Does he maintain open communication channels, is she responsive, and do they usually meet important deadlines? Perhaps unsurprisingly, the interpretation of the strictness of internal deadlines may be culturally dependent and, therefore, this aspect could be particularly challenging. Cultural differences are real. They can enrich the collaboration but also cause friction. So, how can we best approach any negative fallout from culturally diverse research collaborations? Two key aspects come to mind: maintain an open mind as well as a flexible attitude. Be aware of the most pertinent cultural differences between researchers hailing from different countries. You do not need to do extensive research into your potential international collaborator's cultural background, but keeping an open mind, being prepared to respond flexibly, and having a generally heightened awareness of cultural issues will go a long way to resolving the inevitable misunderstandings you will come across [3]. This is also known as developing one's cultural metacognition. And, despite everyone's goodwill and (hopefully) preparation, such instances will continue to occur at irregular intervals. The success of your international joint research project will, therefore, depend on how flexibly you can adjust to changing circumstances.

Other important questions to consider include the following: What is her ability to work well with other team members, particularly at a distance, across national boundaries? Is their work style complementary to yours, thus avoiding unnecessary conflicts? Does including the scientist or team you are targeting lend additional credibility and validity to the project? And if you are a non-native speaker of English (which is, after all, the lingua franca of scientific communication), is your potential partner well versed in writing and/or speaking the language? If all external signs are sufficiently positive to consider taking the next step, reaching agreement on a pilot project or a short-term feasibility study would be a prudent approach to assessing the extent to which your international collaborators' abilities and working practices match your expectations and, subsequently, to potentially establishing a longer-term collaboration.

## Rule 3: Consider Practical Approaches to Establishing the Relationship

As active scientists at any level, from graduate student to senior professor, we encounter numerous opportunities to engage with international colleagues. Although the most obvious openings might arise through interactions at conferences or other meetings, either domestic or international, these are by no means the only suitable or even the most effective means of networking. Many university departments and research institutes run active visitors' programs, often in the form of regular seminar series, which offer an ideal environment in which to meet and get to know international visitors. Conversely, you may have the opportunity to visit foreign institutes. If so, consider in advance to whom you would like to talk and whether you might have common interests that could potentially lead to joint research. And if you have moved to a new institution, do not forget the links you made with colleagues at your previous institute, particularly those who could offer complementary expertise that would benefit joint projects. In general, it is important to proactively pursue any collaborative opportunities.

In many cases, junior scientists will benefit from introductions made by their research supervisor or another local colleague, either in person or by email. And you could, of course, also approach a potential international collaborator yourself, but keep in mind that many people are very busy. It pays, therefore, to make your introduction interesting to your potential collaborator. In addition to asking for their time and/or resources, the most productive collaborations are established if you can offer a substantial benefit in return. This could be in the form of your time (particularly if you can take a leading role in the proposed project); your group's resources, expertise, and external collaborations; or even additional introductions. In today's interconnected world, networking is more important than ever, so by offering such an opportunity to your potential foreign collaborator, you may well have the edge!

## Rule 4: Define the Type of Collaboration You Want to Pursue

You may have considered setting up simple student exchanges, possibly through student co-supervision, but international collaborative projects often offer a multitude of additional opportunities to advance the overall research goals. It thus pays to consider, at an early stage, which type of collaboration you might want to pursue. There is no single answer to this important question, since this depends on the nature of the research question(s) to be tackled, the purpose and/or scope of the study, the extent and nature of the expertise you may require, administrative regulations or restrictions of the institutions involved, preferences of funding agencies, and possibly previous experience with your potential collaborators.

What is the overarching goal of the collaboration that cannot be achieved regionally or even domestically? Is it simply to provide interactions among researchers with different expertise, increased access to resources, or enhanced credibility? Are you after fresh perspectives to avoid "academic inbreeding," or would you like to expand your research network? It is advisable to start relatively small, not least because larger international research collaborations often require substantial administrative resources, e.g., project management and financial reporting. You should be very confident that your international partners are indeed well matched before pursuing such large-scale opportunities. Many funding agencies, including national research councils and some nonprofit organizations, offer competitively awarded seed money for pilot projects or feasibility studies to explore whether the proposed international partners are indeed suitable to pursue more substantial research questions through subsequent joint efforts. Consider this your first step: short-term, fairly straightforward pilot projects are a great and low-risk way to get to know your international partner better and figure out whether you can work together productively and successfully.

## Rule 5: Clearly Define the Main Goals and Expected Outcomes

The golden rule in any collaborative context is to be as specific as you can be about the project's goals during its development phase. Where international collaborations are concerned, it is particularly important to consider the conditions for success, and you should plan to evaluate all aspects of the collaboration rigorously. This is particularly important in the context of larger international collaborations, which are often closely audited externally, but you should adopt a positive attitude with respect to evaluation, by default. Clearly define the team members' roles and responsibilities, and be prepared to offer significant time commitment yourself.

Fostering a high level of cooperative teamwork takes time and effort, particularly in international collaborations where cultural and language differences may prove challenging. Developing trust, collegiality, and a sense of fairness and accountability are at the basis of any successful research collaboration [2,4], irrespective of the team's geographical distribution.

## Rule 6: Be Aware of the Most Important Obstacles to Establishing the Relationship

While establishing regional or domestic collaborations may already be challenging for a variety of mundane reasons, these difficulties might be amplified in an international context. For instance, conflicting research paradigms in different national settings, disagreements on conventions or standards of practice, as well as a lack of compliance with international research protocols may all affect the integrity of the joint research project. In addition, collaborators may not share the same professional jargon, or even speak the same working language sufficiently proficiently. One should certainly also be aware of cultural differences, even for collaborations between presumably similar research partners (such as scientists in the United States and the United Kingdom). While present-day electronic communication has made joint research across national borders viable and often highly successful, face-to-face meetings are arguably the most important vehicle by which to establish personal relationships with your international partner(s). Nothing can beat face-to-face discussions, ideally occurring at an early stage of your joint project, to overcome mundane obstacles such as a language barrier and cultural differences. This applies to both the workplace and social settings. Do not underestimate the importance of getting to know your international partner in a relaxed atmosphere; the best research collaborations are built on excellent personal relationships!

Differing opinions on the goals of and procedures pertaining to the research project, perhaps driven by such cultural differences or caused by disagreements regarding sharing time, work, data, or resources, may inadvertently affect the research collaboration. This particularly applies to awarding co-authorship on any publications resulting from the collaboration, as well as to proper attribution of credit for any of the team's achievements. The most important keywords of relevance to international teamwork are transparency, openness, and careful planning. Above all, be prepared to be flexible while retaining ownership of the research collaboration.

## Rule 7: Discuss Dissemination Policies as well as Intellectual Property Rights at an Early Stage

In some research environments, principal investigators commonly include all members of their research group as co-authors, whereas this may be frowned upon elsewhere. This underscores the importance of reaching agreement on dissemination policies of the research outcomes as well as on authorship criteria. Ideally, this should be discussed at an early stage, well before disagreements may become an obstacle to the successful pursuit of the collaboration. Depending on your research focus and the expected outcomes, this is also the time to discuss and clarify the collaboration's policies on commercialization and intellectual property rights. Note, however, that you will most likely be bound by the regulations established by your institution in this regard, so familiarize yourself with these boundary conditions.

Both the International Committee of Medical Journal Editors and the American Psychological Association have established well-regarded authorship criteria that are commonly adhered to by many international peer-reviewed journals, across disciplines [5,6]. It is recommended to consider the acceptable criteria for inclusion as (co-)author, the standard of acceptability regarding format and content of disseminated results, as well as the proper allocation of credit.

Establishing an effective dissemination policy requires reaching agreement on who will be authorized to speak on behalf of the collaboration, the target audience [7], and the nature and details of the information to be shared, both internally and externally [2]. The most important keywords in this context are trust and collegiality.

## Rule 8: Consider and Clarify the Extent to Which You Are Prepared to Share Resources

You should expect that senior researchers at high-profile international research institutes, particularly those possessing unique expertise, are busy scientists and project managers who receive numerous requests to embark on new collaborative projects every year. Your request will most likely fall on deaf ears if you do not offer anything in return, thus showing that you are a serious contender for your potential collaborator's time and resources. It is, therefore, of the utmost importance to consider what you could offer your potential partner so that you will hit the ground running. Your contribution may include earmarked funding, personnel (such as shared students or postdoctoral researchers), proprietary data or analysis tools, access to specialized equipment, and/or novel ideas.

However, do not be disappointed if your request is turned down. There are numerous reasons why a potential collaborator may not be able to share their resources, including a need to protect preliminary work from criticism, claims related to discovery and priority (e.g., patents), intellectual property issues, safeguarding institutional or local investments, or confidential information (such as that related to peer review, human subjects, and medical data or in private, military, or forensic research [2]).

## Rule 9: Avoid Conflicts of Interest

While international collaborations can be highly rewarding, one should be acutely aware of the myriad conflicts of interest that might arise. In addition to issues related to (co-)authorship and intellectual property rights, the exchange of (medical) data between collaborators and countries can be a significant hurdle, and any data exchange policies need to be clarified at an early stage. In addition, some high-technology components are subject to US export controls under the International Traffic in Arms Regulations [8].

Your most important consideration should be to protect your reputation and your research integrity [9]. Full disclosure of any potential conflicts of interest is recommended, to funding agencies, institutions, collaborators, as well as journal editors. Avoid the appearance of biases, which may occur if such conflicts of interest come to the surface at a later stage. This could compromise the credibility of your study, even if nothing untoward happened. You could face claims of professional misconduct and, in extreme cases, your papers may be retracted.

## Rule 10: Be Aware of Potential Funding Opportunities

Different types of research require vastly different amounts and types of funding. Costs arising from research in an international context may include purchases of materials, travel expenses, publication and other dissemination charges, and personnel costs such as salaries of postdoctoral researchers or exchange students. Which of these costs apply to your collaboration depends on its nature. They can range from pilot funding to explore the viability of establishing a bilateral or multilateral collaboration to funding for exchanges of personnel (either as short- or long-term research projects) and large-scale multinational research projects, funded by the likes of the European Union (EU). Each of these will come with its own requirements and pressures.

In most cases, your main national funding agency is your first port of call. Standard research grants usually include an allowance for international expenditure and exchanges. Specific bilateral projects may be announced on an annual basis, while thematic calls for applications may include foreign investigators. Alternatively, consider submitting targeted funding applications to private, nonprofit foundations. These often focus on specific science, applications, or themes, and your joint research may well fit their briefs. Some of these foundations operate internationally and often target specific countries or regions, such as those in the developing world [10]. Your institution's Research or International Offices are good entry points to start exploring these opportunities.

Finally, although there are many international or bilateral funding opportunities available for which your project may be eligible, it is not always easy to distinguish the wheat from the chaff. In the absence of a master list of funders, it is highly recommended to peruse the opportunities offered by various national research councils—many of which have established a presence well beyond their home nations' borders, such as the British Council, the Alliance Française, the Humboldt Stiftung, and many others—as well as the EU's offerings, which include the Marie Skłodowska-Curie actions and well-maintained lists of funding opportunities aimed at researchers from specific regions [11], including North America, China, and Japan, among others. And do not forget to check the websites of the embassy of the country where your potential research collaborator is based. You may be surprised by what the dedicated teams of science and technology councilors could offer you, although their deadlines may not come around regularly.

## Acknowledgments

In preparing the workshop materials on which this editorial is based, I have particularly benefited from the online tutorial modules on responsible conduct in collaborative research by Northern Illinois University's Faculty Development and Instructional Design Center.

## References

1. Adams J (2012) Collaborations: The rise of research networks. Nature 490 (issue 7420): 335–336. doi: 10.1038/490335a PMID: 23075965
2. Faculty Development and Instructional Design Center, Northern Illinois University (2005) Responsible conduct in collaborative research. http://ori.hhs.gov/education/products/niu_collabresearch/index.html. Accessed 12 December 2014.
3. Blanding M (2012) Collaborating Across Cultures. HBS Working Knowledge. http://hbswk.hbs.edu/item/6687.html. Accessed 18 February 2015.
4. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3(3): e44. PMID: 17397252
5. International Committee of Medical Journal Editors (2013) Defining the role of authors and contributors. http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html. Accessed 12 December 2014.
6. American Psychological Association (2014) Publication Practices & Responsible Authorship. http://www.apa.org/research/responsible/publication/. Accessed 14 December 2014.
7. de Grijs R (2009) Work with the media: Think audience! http://www.southernscience.co.za/astronomystars/media01.php. Accessed 12 December 2014.
8. US State Department (2014) International Traffic in Arms Regulations, 2014 update. https://www.pmddtc.state.gov/regulations_laws/itar.html. Accessed 12 December 2014.
9. Bourne PE, Barbour V (2011) Ten simple rules for building and maintaining a scientific reputation. PLoS Comput Biol 7(6): e1002108. doi: 10.1371/journal.pcbi.1002108 PMID: 21738465

10. Moreno E., Gutiérrez J-M (2008) Ten simple rules for aspiring scientists in a low-income country. PLoS Comput Biol 4(5): e1000024.

11. European Commission (2014) EURAXESS Researchers in Motion. http://ec.europa.eu/euraxess/index.cfm/links/index. Accessed 12 December 2014.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for a Successful Cross-Disciplinary Collaboration

**Bernhard Knapp[1]*, Rémi Bardenet[1], Miguel O. Bernabeu[2,3], Rafel Bordas[4], Maria Bruna[4,5], Ben Calderhead[6], Jonathan Cooper[4], Alexander G. Fletcher[5], Derek Groen[2], Bram Kuijper[3,7], Joanna Lewis[3], Greg McInerny[4], Timo Minssen[8,9,10], James Osborne[4], Verena Paulitschke[11], Joe Pitt-Francis[4], Jelena Todoric[12], Christian A. Yates[13], David Gavaghan[4], Charlotte M. Deane[1]**

**1** Department of Statistics, University of Oxford, Oxford, United Kingdom, **2** Centre for Computational Science, Department of Chemistry, University College London, London, United Kingdom, **3** Centre for Mathematics, Physics and Engineering in the Life Sciences and Experimental Biology, University College London, London, United Kingdom, **4** Department of Computer Science, University of Oxford, Oxford, United Kingdom, **5** Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Oxford, United Kingdom, **6** Department of Mathematics, Imperial College London, London, United Kingdom, **7** Department of Genetics, Evolution and Environment, University College London, London, United Kingdom, **8** Centre for Information & Innovation Law, University of Copenhagen, Copenhagen, Denmark, **9** Guest Scholar, Harvard Law School, Cambridge, Massachusetts, United States of America, **10** Visiting Research Fellow, University of Oxford, Oxford, United Kingdom, **11** Department of Dermatology, Medical University of Vienna, Vienna, Austria, **12** Laboratory of Gene Regulation and Signal Transduction, Departments of Pharmacology and Pathology, University of California, California, San Diego, United States of America, **13** Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

* bernhard.knapp@stats.ox.ac.uk



OPEN ACCESS

**Citation:** Knapp B, Bardenet R, Bernabeu MO, Bordas R, Bruna M, Calderhead B, et al. (2015) Ten Simple Rules for a Successful Cross-Disciplinary Collaboration. PLoS Comput Biol 11(4): e1004214. doi:10.1371/journal.pcbi.1004214

**Published:** April 30, 2015



**Funding:** The 2020 Science programme is funded through the EPSRC Cross-Disciplinary Interface Programme (grant number EP/I017909/1). The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Cross-disciplinary collaborations have become an increasingly important part of science. They are seen as key if we are to find solutions to pressing, global-scale societal challenges, including green technologies, sustainable food production, and drug development. Regulators and policy-makers have realized the power of such collaborations, for example, in the 80 billion Euro "Horizon 2020" EU Framework Programme for Research and Innovation. This programme puts special emphasis on "breaking down barriers to create a genuine single market for knowledge, research and innovation" (http://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020).

Cross-disciplinary collaborations are key to all partners in computational biology. On the one hand, for scientists working in theoretical fields such as computer science, mathematics, or statistics, validation of predictions against experimental data is of the utmost importance. On the other hand, experimentalists, such as molecular biologists, geneticists, or clinicians, often want to reduce the number of experiments needed to achieve a certain scientific aim, to obtain insight into processes that are inaccessible using current experimental techniques, or to handle large volumes of data, which are far beyond any human analysis skills.

The synergistic and skilfulcombining ofdifferent disciplines can achieve insight beyond current borders and thereby generate novel solutions to complex problems. The combination of methods and data from different fields can achieve more than the sum of the individual parts could do alone. This applies not only to computational biology but also tomany other academic disciplines.

Initiating and successfully maintaining cross-disciplinary collaborations can be challenging but highly rewarding. In a previous publication in this series, ten simple rules for a successful collaboration were proposed [1]. In the present guide, we go one step further and focus on the specific challenges associated with cross-disciplinary research, from the perspective of the theoretician in particular. As research fellows of the 2020 Science project (http://www.2020science.net) and collaboration partners, we bring broad experience of developing interdisciplinary collaborations. We intend this guide to be for early career computational researchers as well as more senior scientists who are entering a cross-disciplinary setting for the first time. We describe the key benefits, as well as some possible pitfalls, arising from collaborations between scientists with very different backgrounds.

## Rule 1: Enjoy Entering a Completely New Field of Research

Collaborating with scientists from other disciplines is an opportunity to learn about cutting-edge science directly from experts. Make the most of being the novice. No one expects you to know everything about the new field. In particular, there is no pressure to understand everything immediately, so ask the "stupid" questions. Demonstrating your interest and enthusiasm is of much higher value than pretending to know everything already. An interested audience makes information sharing much easier for all partners in a collaboration.

You should prepare for a deluge of new ideas and approaches. It is a good practice to read relevant textbooks and review papers, which your collaborators should be able to recommend, in order to quickly grasp the vocabulary (see Rule 3) and key ideas of the new field. This will make it easier for you to establish a common parlance between you and your collaborators, and allow you to build from there.

You should try to discuss your work with a range of scientists from complementary fields. As well as getting feedback, this can help you identify new collaborative opportunities. Remember that contacts that do not lead directly to collaborations can still prove useful later in your career.

## Rule 2: Go to the Wet Lab

It is vitally important to understand where specific data sets come from. Just like mathematical and computational models, experiments have their own in-built assumptions, strengths, and weaknesses that you need to understand. What was the exact process of data collection? How many experiments can be performed in a given timeframe and how much do they cost? What were the constraints that led to the design of the experiments—how will you include this in your interpretation? If you plan to use the resulting data for model calibration or parameter fitting then try to obtain sufficient information to reproduce the experiment in silico. Papers in different domains have different perspectives and might not contain the data you are looking for in sufficient detail. Visiting the lab in person is often the most efficient way to get the information you need. A good understanding of the experimental setup might also suggest appropriate testcases for the computational studies. Try to talk to both the junior and senior scientists in the lab as they may give you different perspectives.

There are social, as well as scientific, reasons for understanding life in the wet lab. As a computational scientist, it is easy to underestimate the commitment and resources necessary to acquire experimental data (see rule 4). Visiting a lab, and taking an interest in data collection, is a way of acknowledging your colleagues' effort and the value of their data and expertise.

## Rule 3: Different Fields Have Different Terminologies: Learn the Language

Science is full of subcultures using diverse and evolving jargon. Forming a successful cross-disciplinary relationship requires that you fully understand your collaboration partner. From classification schemes and methods to journals and research philosophy, it can be hard enough keeping up with developments in your own field, let alone others. For instance, neologisms can be ubiquitous in computational and biological sciences, where new terminology continually emerges from new methods, tools, and knowledge. Learn the other field's jargon early on in the collaboration and ask basic questions about the meanings of words.

For example:

- Ambiguity: "Model" is probably the most ambiguous word in science. Mathematical, statistical, experimental, observational, theoretical, computational, analytical, verbal, legal, mental, graphical, geometrical, structural, and workflow models all have different meanings. Almost every field will have its own interpretation of "model" and the semantics differ significantly.
- Synonyms: For example, removing entities above and below certain thresholds is termed "positive and negative selection" in immunology, while it is called "band-pass filter" in signal transduction.

Context often matters, so try to understand nuances in the use of terms. It can be beneficial to build up a technical glossary. Evaluate your understanding by presenting it back to new colleagues and observe where your rudimentary understanding needs more work.

Finally, agree on a joint nomenclature with your collaborators early in the project. Write equations and code in a consistent manner, standardise data formats, and use consistent style schemes in figures. Then talk through your outputs to discuss your collaborators' understanding and involvement. A good relationship is based on mutually understandable communication.

## Rule 4: Different Fields Move at Different Speeds: Do not Become Impatient

A huge variety of cultures and expectations regarding research and subsequent publication exist in different scientific disciplines. However, these differences can lead to stress when embarking upon multidisciplinary collaborations, unless they are acknowledged and effectively communicated at an early stage. It is important to accept the different pace of different fields, communicate well, and be patient.

Research in experimental biology, for example, often involves long and arduous experiments, taking perhaps months or even years to complete. Animals or tissues may need to be grown, and weekends or nights spent in the lab tending to cell cultures and repeating experiments may be necessary. Some projects generate publications and co-authorships several years after a theoretician may have actually performed their in silico contribution to the work. Vice versa, computational aspects often involve more than simply pressing a button and computational resources may be limited.

Do not make assumptions about how hard fellow collaborators are working based on how long they take to get back to you with results. Here, communication is of particular importance.

Similarly, journals in different disciplines might have different periods of time from submission to publication. This can have knock-on effects when demonstrating your research output (see rule 5).

Early communication of how long your part of the work is likely to take and why this amount of time is needed will help your collaboration to run more smoothly.

## Rule 5: Different Fields Have Different Reward Models: Know What You Can Expect

It is important to recognise that the publication culture in the life sciences, and in experimental biology particularly, differs from that of the theoretical sciences. Such differences can include:

- Publication speed varies greatly. In experimental biology, publishing often takes several years, while certain theoretical papers can be published in a much shorter timescale (see also rule 4).
- Metrics, such as the impact factor (IF), are used by many organisations to evaluate your research [2,3]. Be aware that different fields have different impact factor scales. The journal impact factors mainly depend on the average length of reference lists in the field. For example, a journal with an impact factor of 3 in mathematics (295 journals, median IF 0.57, maximum IF 3.57) might be more prestigious than a journal with an impact factor of 30 in cell biology (184 journals; median IF 3.2; maximum IF 37.16) (based on Journal Citation Reports of Thomson Reuters, version 2012).
- In some fields, such as information technology, it can be the norm to publish new research in peer-reviewed conference proceedings instead of journals.
- The preferred ordering of authors on a manuscript may also depend strongly on the academic environment. The first author might be the scientist who contributed most or whose surname comes first alphabetically. The last author may be the principal investigator, the author who contributed least, or the author with the last surname alphabetically. The corresponding author can be seen as "in charge of the paper," the principal investigator, or the person who volunteered for dealing with the correspondence.
- In some areas of biology, large consortia of authors are needed to conduct research. In some theoretical fields, people tend to publish with fewer authors. Thus, the definition of a "significant" contribution to a manuscript might differ markedly.

It is important to be aware of these differences. Make sure that you discuss all these issues early with your collaborators to avoid misunderstandings and frustration. A well-planned publication strategy is fundamental in order to fulfil everyone's expectations and accommodate the potential mismatch of timescales of theoretical/experimental work (see rule 4).

Does your field and that of your partner value less frequent, higher impact publications or a series of smaller publications? One option is to start with methodological papers (both theoretical and experimental) while final publications describing the major breakthrough and how all the components are brought together could follow. Early methodological papers should already highlight the benefits of collaborating, e.g., theoretical work with experimentally sound assumptions and parameters, experimental work with solid data analytics. It is advisable to design initial publications without forgetting the greater scope of the collaboration. However, be aware that preceding papers might weaken your main publication if they anticipate parts of the results.

## Rule 6: What Different Fields Mean by "Data"

Be prepared that scientists with experimental backgrounds might not have the same structured view on data and terminologies (see also rule 3) as you have. For scientists with a background in computer science, the lowest level of data organisation might be a spreadsheet where each column and row is well defined. For scientists with non-technical backgrounds, such a spreadsheet might represent the highest form of data organisation.

Whenever possible, ask your collaboration partner for a standardized data format. Good guides on how to share data can be found in [4,5]. Always favour electronic forms of data and always keep a copy of the original file. You might also consider writing minutes about meetings and data specifications to avoid later misunderstandings.

Do not blindly trust experimental data. Always perform "sanity checks" on the data you receive (graphs, frequency tables, mutually exclusive data, unnatural distributions, etc.). This will help you to see if you have interpreted the data correctly and allow you to ask questions if you have any doubts.

## Rule 7: Assess the Advantages and Disadvantages of Service Work

Theoretical scientists can be a huge asset to multidisciplinary projects by providing experimentalists with computational tools to gather data, predictive methods, and advanced statistical modelling. Theoretical scientists often do substantial amounts of "service work" in a collaborative project, for example, by maintaining computer infrastructure and databases, keeping their in-house code base up-to-date, and statistically analysing data.

Service work is often an excellent way to establish a collaboration, get the partners to trust in your ability and expertise, and learn enough about other disciplines to start making direct contributions, whilst, at the same time, co-authoring high-quality publications. Service work will show that you take the collaboration seriously and help you to establish a reputation as a reliable and analytically keen scientist who delivers fast, structured, and correct results.

Nonetheless, service work is also risky, as it may take more time than you anticipated. Therefore, make sure to evaluate the amount of service work on a regular basis and be clear with your collaborators about what you expect in return before engaging in service work. To gain more insight into the "cost" of service work, keep a record of the amount of time spent on service tasks. This not only prevents your collaborators from treating your contributions lightly but also gives you a clear idea as to whether it is worthwhile to engage in such tasks and/or take on new ones.

Crucial to minimizing the service "'load" is to make it easy to delegate tasks to others. When starting to develop analytical tools for others (whether software or mathematics), always take the perspective that these tools won't primarily be used by you, but by other collaborators. Hence, making your tools user-friendly, for example, by providing illustrative examples and documenting your code extensively [6], is essential.

## Rule 8: Create and Manage Structural Bonds

Cross-disciplinary collaborations require structural bonds between the collaborators. It is only possible to break the silos of scientific disciplines and to become truly cross-disciplinary if a proper framework for scientific exchange is established. This can include regular meetings, workshops, symposia, attendance of each other's group meetings, and co-teaching of courses. However, to keep your collaboration efficient, be careful about imposing too many obligations. While it is often necessary to leave the "comfort zone"of scientific disciplines, it is equally important not to frustrate your collaboration partners with too many details not relevant to their endeavours. Therefore, keep the number of meetings at a reasonable level and set clear agendas.

Moreover, establishing these bonds often requires financial support, which can be achieved in different ways. For the initial setup phase, seed funding schemes maybe a useful resource. On the basis of these initial bonds, applications for larger grants can be submitted collaboratively. Many funding bodies offer special calls for cross-disciplinary research or favour cross-disciplinary proposals for both national and international settings. Examples include the

"Horizon 2020" EU Framework (http://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020), the Human Frontiers in Science Program (http://www.hfsp.org/), the US NSF/BIO–UK BBSRC Lead Agency Pilot (http://www.bbsrc.ac.uk/funding/internationalfunding/nsfbio-lead-agency-pilot.aspx), or internal projects created to achieve inter-faculty cooperation within a university.

Once funding is secured, shared PhD students and postdocs can further strengthen bonds between collaboration partners. Shared supervision rewards with constant knowledge exchange, shared publications, and an interdisciplinary training for all. However, it is also important to protect junior scientists from "getting lost in cross-disciplinary collaboration."In particular, there is a risk that they fall between two stools in their attempts to comply with the expectations and advice of two or more supervisors from completely different fields.

## Rule 9: Recognise When Things Are Not Working Well

Unlike a marriage, collaborations are not necessarily intended to be continuous and permanent. A pragmatic approach can favour both parties. If major problems arise that cannot be solved after a couple of attempts, there are several possible next steps:

1. Ostrich approach: Pretend that nothing has happened and hope that normality will soon be restored. We strongly advise against this tactic as it might lead to more frustration and potential damage to the relationship.
2. Pause: Sometimes, one of the collaborators may find it harder than expected to deliver the agreed results or is overwhelmed with other work-related duties. If your collaborator has trouble with their side of the workload, it is often better to introduce a pause in the collaboration than to exert pressure on them. Deliberate pauses take pressure off your collaborations, and can save potential frustration on your side as well.
3. Search for alternatives: Collaborations do not need to be exclusive. If you prioritise the collective work more highly than your collaborator, you might consider establishing a fresh collaboration with someone else who also values the work more. But be aware of your actions and the consequences they might have on your current collaboration.
4. End a collaboration: If a collaboration has become unworkable or reached a natural terminus, it may be best to make a clean cut and end the collaboration once and for all.

Many failures in collaborations could have been avoided by an early, proactive approach on arising problems. In cross-disciplinary settings, it could just be a problem of understanding (rule 3), impatience (rule 4), or lack of reward (rule 5). If you decide to end a collaboration, make sure to keep a working relationship with your former collaborator. This will allow you to properly handle existing structural bonds (rule 8) and allow you to potentially initiate other collaborations with the same partner.

## Rule 10: Be Synergistic

Probably the most important quality of collaborations is the mutual gain that emerges. This usually works best when scientists with different, but complementary, skills decide to work together. One example might be the application of a novel, high-throughput computer algorithm to a vast quantity of experimental data. While the algorithm by itself might be brilliant, and publishable by benchmarking it on a publicly available dataset, it will shine even more when applied to a huge, unpublished dataset. At the same time, the huge dataset could be analysed by standard, semi-manual methods. This might also be publishable, but would take a long time and essential insights may be missed, which the novel algorithm would have delivered. Only by combining contributions from both sides does the work become more than the sum of its parts, achieving useful things for each partner and enabling insights that would not have been possible for either alone.

Initiating and nurturing successfully synergistic relationships is an important and valuable skill. The success of an interdisciplinary collaboration depends on any number of factors, including which partner approached which and over what timescale each party envisages collaboration. One very important property for a long-term, effective, and mutually beneficial relationship is that both sides should feel they are winners during and, especially, after a project (see also rule 5). An outstanding book on how such win/win situations can be achieved is Fisher, Ury, and Patton's *Getting to Yes*: *Negotiating Agreement Without Giving In* [7]: their advice includes inventing options for mutual gain, making sure to always give enough credit to partners, and caring for their interests as you would for your own. Then you will establish a truly successful and synergistic collaboration.

## References


1. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3: e44. PMID: 17397252
2. Eisen JA, Maccallum CJ, Neylon C (2013) Expert failure: re-evaluating research assessment. PLoS Biol 11: e1001677. doi: 10.1371/journal.pbio.1001677 PMID: 24115910
3. Eyre-Walker A, Stoletzki N (2013) The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. PLoS Biol 11: e1001675. doi: 10.1371/journal.pbio.1001675 PMID: 24115908
4. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. (2014) Ten simple rules for the care and feeding of scientific data. PLoS Comput Biol 10: e1003542. doi: 10.1371/journal.pcbi.1003542 PMID: 24763340
5. White EP, Baldridge E, Brym ZT, Locey KJ, McGlinn DJ, Supp SR (2014) Nine simple ways to make it easier to (re)use your data. PeerJ PrePrints 1:e7v2.
6. Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, Dalchau N, et al. (2014) Ten Simple Rules for Effective Computational Research. PLoS Comput Biol 10: e1003506. doi: 10.1371/journal.pcbi.1003506 PMID: 24675742
7. Fisher R., Ury W. L., and Patton B. (2011) Getting to Yes: Negotiating Agreement Without Giving In. Penguin Books.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Cultivating Open Science and Collaborative R&D

**Hassan Masum[1]*, Aarthi Rao[2], Benjamin M. Good[3], Matthew H. Todd[4], Aled M. Edwards[5], Leslie Chan[6], Barry A. Bunin[7], Andrew I. Su[3], Zakir Thomas[8], Philip E. Bourne[9]**

**1** Waterloo Institute for Complexity and Innovation, Waterloo, Ontario, Canada, **2** Results for Development Institute, Washington, D.C., United States of America, **3** Department of Molecular and Experimental Medicine, Scripps Research Institute, La Jolla, California, United States of America, **4** School of Chemistry, University of Sydney, Sydney, New South Wales, Australia, **5** Structural Genomics Consortium, University of Toronto, Toronto, Ontario, Canada, **6** Department of Social Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada, **7** Collaborative Drug Discovery, Burlingame, California, United States of America, **8** Council of Scientific and Industrial Research, New Delhi, India, **9** Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California, United States of America

How can we address the complexity and cost of applying science to societal challenges?

Open science and collaborative R&D may help [1–3]. *Open science* has been described as "a research accelerator" [4]. Open science implies open access [5] but goes beyond it: "Imagine a connected online web of scientific knowledge that integrates and connects data, computer code, chains of scientific reasoning, descriptions of open problems, and beyond …. tightly integrated with a scientific social web that directs scientists' attention where it is most valuable, releasing enormous collaborative potential." [1].

Open science and collaborative approaches are often described as *open source*, by analogy with open-source software such as the operating system Linux which powers Google and Amazon—collaboratively created software which is free to use and adapt, and popular for Internet infrastructure and scientific research [6,7]. However, this use of "open source" is unclear. Some people use "open source" when a project's results are free to use, others when a project's process is highly collaborative [4].

It is clearer to classify open source and open science within a broader class of *collaborative R&D*, which can be defined as scalable collaboration (usually enabled by information technology) across organizational boundaries to solve R&D challenges [8].

Many approaches to open science and collaborative R&D have been tried [1,9]. The Gene Wiki has created over 10,000 Wikipedia articles, and aims to provide one for every notable human gene [10]. The crowdsourcing platform InnoCentive has reportedly facilitated solutions to roughly half of the thousands of technical problems posed on the site, including many in life sciences such as the $1 million ALS Biomarker Prize [11]. Other examples include prizes (X-Prize [12]), scientific games (FoldIt [13]), and licensing schemes inspired by open-source software (BIOS [14]).

Collaborative R&D approaches vary in openness [15]. In some approaches, the R&D process and outputs are open to all—for example, open-science projects like the Gene Wiki described above. In other approaches which demonstrate what might be called *controlled collaboration*, there are strong controls on who contributes and benefits—for example, computational platforms like Collaborative Drug Discovery or InnoCentive that support both commercial and nonprofit research [9,11].

Collaborative approaches can unleash innovation from unforeseen sources, as with crowdsourcing health technologies [11–13,16]. They may help in global challenges like drug development [17], as with India's OSDD (Open Source Drug Discovery) project that recruited over 7,000 volunteers [16] and an open-source drug synthesis project that improved an existing drug without increasing its cost [18].

If you want to apply open science and collaborative R&D, what principles are useful? We suggest Ten Simple Rules for Cultivating Open Science and Collaborative R&D. We also offer eight conversational interviews exploring life experiences that led to these rules (Box 1).

## Rule 1: Get the Incentives Right—Learn from the Past

Why should contributors take part in your project? Learn from incentives that have worked in mass collaborations and open-source software, such as reputation building, enjoyment, cooperatively solving interesting problems that are too hard to do alone, and jointly developing tools that benefit all developers [6,7,19]. Organizational incentives can include lowering costs, tapping external innovation, implementing novel business models such as selling complementary services, and jointly competing for public admiration or grant funding. Altruism can motivate collaboration, but frequently it is not the main reason [9]. With this in mind, align individual incentives with collective benefit [1]. Look to past and present precompetitive collaborations for ways to address intellectual property and competitive concerns [3]. Share attribution with contributors so they can advance their goals and demonstrate their capabilities.

## Rule 2: Make Your Controlled Collaborations Win-Win-Win

Perhaps completely open science seems unsuitable to you, if for example you are engaged in market-driven R&D that must recoup investments. There are ways to

**Citation:** Masum H, Rao A, Good BM, Todd MH, Edwards AM, et al. (2013) Ten Simple Rules for Cultivating Open Science and Collaborative R&D. PLoS Comput Biol 9(9): e1003244. doi:10.1371/journal.pcbi.1003244

**Editor:** Fran Lewitter, Whitehead Institute, United States of America

**Published** September 26, 2013



**Funding:** This article was made possible by support to HM and AR from a grant by the Bill & Melinda Gates Foundation to the Results for Development Institute. The funders had no role in the preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: hassan.masum@utoronto.ca
Philip E. Bourne is Founding Editor-in-Chief of *PLOS Computational Biology*.

### Box 1. Conversations on Open Science and Collaborative R&D

Many commentators have considered challenges in translating open science and collaborative methods to biomedical research [2–4,9,17,20,24,26,28,29]. How can protecting intellectual property be balanced with freeing researchers to build on previous knowledge? If R&D results are collaboratively created and freely available, who will take responsibility for costly clinical trials and quality control? What will be the Linux of open-source R&D?

To explore such challenges and convey life experiences in biomedical open science and collaborative R&D, we offer eight conversational interviews by the first author of this article as supplementary material. The conversations were done on behalf of the Results for Development Institute and are with:

- **Alph Bingham**, cofounder of InnoCentive (Text S1)
- **Barry Bunin**, CEO of Collaborative Drug Discovery (Text S2)
- **Leslie Chan**, open access pioneer and director of Bioline International (Text S3)
- **Aled Edwards**, director of the Structural Genomics Consortium (Text S4)
- **Benjamin Good**, coleader of the Gene Wiki initiative (Text S5)
- **Bernard Munos**, pharmaceutical innovation thought leader (Text S6)
- **Zakir Thomas**, director of India's Open Source Drug Discovery (OSDD) project (Text S7)
- **Matt Todd**, open science and drug development pioneer (Text S8)

benefit from open science and collaborative methods while retaining appropriate controls and the opportunity to provide public benefit. You, your partners, and the public can all benefit—a win-win-win situation. You might use computational platforms to supercharge information sharing with selected partners, including public-benefit initiatives that match your mission [9]. You might use crowdsourcing to overcome roadblocks by opening up chosen parts of your R&D process to new innovators [11]. Or you might make public selected data or software tools, exporting them to the open-source realm to gain from goodwill or quality improvement [3]. Sharing can make both business and social sense, whether in implementing open standards, collaborating precompetitively, or reducing duplication of effort [20]. Keep an eye open for opportunities to "do well by doing good" by structuring initiatives for private and public benefit [21]. Collaborative approaches can benefit both public and private sectors in collaborating across competitive boundaries, connecting problems with problem solvers, and cultivating a knowledge commons [1,9].

## Rule 3: Understand What Works—and What Doesn't

You can save yourself frustration by not using an unsuitable collaborative method, be it a Wiki without an audience or a crowdsourced research challenge without focus [8]. Consider questions like: have you learned from others who have tried the method? Do you understand when the method fails, and what is necessary for it to work? Is there a good match between the method and your goals? Are you contributing your experiences and interesting failures back to the community, thus demonstrating thought leadership? If you are interested in more effective knowledge sharing, consider low-budget opportunities such as starting an online Q&A site about open science or collaborative R&D using a platform like StackExchange. There are also opportunities to help evaluate what really works—moving beyond anecdotal evidence to case studies and metrics.

## Rule 4: Lead as a Coach, Not a CEO

The command-and-control style doesn't work well with contributors from diverse organizations, many of whom may be volunteers [22]. And as has been said of Linus Torvalds, the founder of the open-source operating system Linux, "Linus doesn't scale": leaders of mass collaborations can become bottlenecks unless they encourage distributed workflows and leadership [7]. Be flexible about management (but strict about quality). Check your ego at the door—you're playing a team game and will be stronger when others want to contribute. Participants will feel more motivated if their contribution enriches a joint resource rather than just the leader. Can you give up exclusive ownership and credit to achieve with others what you cannot achieve alone?

## Rule 5: Diversify Your Contributors

A powerful aspect of collaborative R&D is the potential diversity of the community—including students [16], patients [23], gamers [13], and researchers from lesser-known countries or institutions. You can use open science to attract diverse contributors by lowering barriers to participation, publicly tackling audacious challenges (see Rule 8), and making collaboration fun. Consider open licensing terms and joint or public ownership of selected outcomes to broaden your participant base [14,15,21,24]. Encourage all community members to find ways to contribute that suit their abilities and inclinations. Can you reach past your usual partners, and make it easy for others to get up to speed with what you're doing? Are there opportunities for "citizen science," perhaps through organizing many microcontributions [1,13]?

## Rule 6: Diversify Your Customers

Can you engage the broadest possible base as beneficiaries? The science that you do in the open spreads its benefits widely, and that can attract unexpected accolades and collaborators [1,4]. Productively involving stakeholders can inform your research—for example, through participatory research strategies involving the people your efforts are meant to help [25]. Contributing to collaborative initiatives targeting human development challenges can motivate your team, and potentially lead to innovations that are transferable to for-profit markets. Neglected disease R&D is a case in point which seems particularly suitable for collaborative pilot projects, given its lower profits, humanitarian appeal, and need for new methods [26]. If your work is commercially driven, consider humanitarian licensing approaches that encourage nonprofit applications by others to poorer demographics [2,21].

## Rule 7: Don't Reinvent the Wheel

The more you can use what already exists, the greater your effectiveness will be. Are there lab and computational resources that could be used when otherwise idle? Can you find people already working on elements of your problem, and organize their collective work? Before starting a new initiative, have you explored and considered joining existing ones? Piggybacking on active efforts eases prototyping and gathering enthusiastic initial users. Build on the cumulative stockpile of past open initiatives (see Rules 1 and 3).

## Rule 8: Think Big

For projects hoping to harness the power of mass collaboration, a major challenge can be attracting a large community of contributors. Many of the best mass collaborations orient around seemingly audacious goals like: "build a free encyclopedia of all the world's knowledge" (Wikipedia), "develop a review article for every human gene" (Gene Wiki), and "build a new operating system" (Linux). Establishing a driving, high-level purpose will help spread the idea of your project and motivate people to come have a look and see what they can do. Be ready to scale with success.

## Rule 9: Encourage Supportive Policies and Tools

Can you cultivate open science and collaborative R&D by helping to make them part of "standard operating procedure"? For example, can you encourage institutional data sharing [24]? Can you build a profiling platform of collaborative initiatives, summarizing what they have achieved and what types of collaborators they are seeking? Do you have opportunities to adopt appropriate policies in your own organization or field? A case study to learn from is the spread of open access from wishful thinking to widespread fact [5].

## Rule 10: Grow the Commons

As intellectual property debates illustrate, there are legitimate differences of opinion on how best to motivate innovators' investments to generate new knowledge [21,26]. But in the long run, sharing more knowledge and tools boosts both for-profit and nonprofit research [2,3]. This growing shared resource of knowledge and tools—"the commons"—is the product of centuries of striving. It depends on cumulative win-win-win collaborations spanning organizations, nations, and generations. Can you find ways to advance your interests while remaining part of this larger narrative [1,5,19,27]?

## Supporting Information

**Text S1** A conversation with Alph Bingham, cofounder of InnoCentive.
(PDF)

**Text S2** A conversation with Barry Bunin, CEO of Collaborative Drug Discovery.
(PDF)

**Text S3** A conversation with Leslie Chan, open access pioneer and director of Bioline International.
(PDF)

**Text S4** A conversation with Aled Edwards, director of the Structural Genomics Consortium.
(PDF)

**Text S5** A conversation with Benjamin Good, coleader of the Gene Wiki initiative.
(PDF)

**Text S6** A conversation with Bernard Munos, pharmaceutical innovation thought leader.
(PDF)

**Text S7** A conversation with Zakir Thomas, director of India's Open Source Drug Discovery (OSDD) project.
(PDF)

**Text S8** A conversation with Matt Todd, open science and drug development pioneer.
(PDF)

## Acknowledgments

We thank Jean Arkedis, Robert Hecht, and Paul Wilson for comments on early versions of this article. Our thanks also go to all the colleagues and pioneers who have shared their wisdom on making collaborative R&D work.

## References


1. Nielsen M (2011) Reinventing discovery. Princeton, NJ: Princeton University Press.
2. National Research Council (2011) Designing the microbial research commons: proceedings of an international workshop. Washington, D.C.: The National Academies Press.
3. IOM (Institute of Medicine) (2011) Establishing precompetitive collaborations to stimulate genomics-driven product development: workshop summary. Washington, D.C.: The National Academies Press.
4. Woelfle M, Olliaro P, Todd MH (2011) Open science is a research accelerator. Nat Chem 3: 745–748. doi:10.1038/nchem.1149.
5. PLOS (2013) PLOS collections: open access collection. Available: http://www.ploscollections.org/openaccess. Accessed 25 April 2013.
6. Prlić A, Procter JB (2012) Ten simple rules for the open development of scientific software. PLoS Comput Biol 8: e1002802. doi:10.1371/journal.pcbi.1002802.
7. Fogel K (2013) Producing open source software. Available: http://producingoss.com/en/. Accessed 25 April 2013.
8. Results for Development Institute (2012) Collaborative health R&D primer. Available: http://healthresearchpolicy.org/primer/. Accessed 25 April 2013.
9. Ekins S, Hupcey MAZ, Williams AJ (2011) Collaborative computational technologies for biomedical research. Hoboken, NJ: John Wiley & Sons, Inc.
10. Good BM, Clarke EL, de Alfaro L, Su AI (2011) The Gene Wiki in 2011: community intelligence applied to human gene annotation. Nucleic Acids Res 40: D1255–1261. DOI:10.1093/nar/gkr925.
11. Bingham A, Spradlin D (2011) The open innovation marketplace. Upper Saddle River, NJ: FT Press.
12. Wilson P, Palriwala A (2011) Prizes for global health technologies. Results for Development Institute. Available: http://healthresearchpolicy.org/assessments/prizes-global-health-technologies. Accessed 25 April 2013.
13. Good BM, Su AI (2011) Games with a scientific purpose. Genome Biol 12: 135. doi:10.1186/gb-2011-12-12-135.
14. Jefferson R (2006) Science as social enterprise: the CAMBIA BIOS initiative. Innovations: Technology, Governance, Globalization 1: 13–44. doi:10.1162/itgg.2006.1.4.13.
15. PLOS (2012) HowOpenIsIT? Available: http://www.plos.org/about/open-access/howopenisit/. Accessed 25 April 2013.
16. Vashisht R, Mondal AK, Jain A, Shah A, Vishnoi P, et al. (2012) Crowd sourcing a new paradigm for interactome driven drug target identification in Mycobacterium tuberculosis. PLoS ONE 7: e39808. doi:10.1371/journal.pone.0039808.
17. Munos B, Chin W (2011) How to revive breakthrough innovation in the pharmaceutical industry. Sci Transl Med 3: 89cm16. doi:10.1126/scitranslmed.3002273.
18. Woelfle M, Seerden J-P, de Gooijer J, Pouwer K, Olliaro P, et al. (2011) Resolution of praziquantel. PLoS Negl Trop Dis 5: e1260. doi:10.1371/journal.pntd.0001260.
19. Benkler Y (2011) The penguin and the Leviathan: the triumph of cooperation over self-interest. New York, NY: Crown Business.
20. Norman TC, Bountra C, Edwards AM, Yamamoto KR, Friend SH (2011) Leveraging crowdsourcing to facilitate the discovery of new medicines. Sci Transl Med 3: 88mr1. doi:10.1126/scitranslmed.3002678.
21. Krattiger AF (2007) Intellectual property management in health and agricultural innovation: a handbook of best practices. Oxford, UK; Davis, CA: MIHR, PIPRA.
22. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3: e44. doi:10.1371/journal.pcbi.0030044.
23. Wicks P, Vaughan TE, Massagli MP, Heywood J (2011) Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat Biotechnol 29: 411–414. doi:10.1038/nbt.1837.
24. Dyke SOM, Hubbard TJP (2011) Developing and implementing an institute-wide data sharing policy. Genome Med 3: 60. doi:10.1186/gm276.
25. Holland J (2013) Who counts?: the power of participatory statistics. Rugby, UK: Practical Action Publishing.
26. Masum H, Harris R (2011) Open source for neglected diseases: magic bullet or mirage? Results for Development Institute. Available: http://healthresearchpolicy.org/assessments/open-source-neglected-diseases-magic-bullet-or-mirage. Accessed 25 April 2013.

27. Masum H, Tovey M (2006) Given enough minds…: bridging the ingenuity gap. First Monday 11. doi:10.5210/fm.v11i7.1370.

28. Marden E (2010) Open source drug development: a path to more accessible drugs and diagnostics? Minnesota Journal of Law, Science and Technology 11: 217–266.

29. Årdal C, Røttingen J-A (2012) Open source drug discovery in practice: a case study. PLoS Negl Trop Dis 6: e1827. doi:10.1371/journal.pntd.0001827.

Editorial

# Ten Simple Rules for Getting Involved in Your Scientific Community

**Magali Michaut***

The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada

A scientific community consists of scientists working in a particular field of science and, most importantly, of their relationships and interactions. Beyond the traditional publication of research projects, discussions occurring during conferences, seminars, and even online through social networks or blogs enable ideas to spread more efficiently and are essential for building a lively and dynamic community. Activities such as organizing conferences and workshops, answering questions and discussing scientific ideas online, contributing to a scientific blog, or participating in open source software projects are typically thought of as outside classic research activity. Having scientists involved in those activities, however, is very important for the community to be dynamic and to promote fruitful discussions and collaborations. Scientific associations have an important role in enabling science by bringing people together and giving them a voice. Moreover, being involved in such activities is individually very rewarding because it enables scientists to acquire new skills not typically taught and to expand their network and interactions.

For those reasons, I encourage young scientists to get involved in their scientific community. However, it should be noted that this involvement takes time during which you are not directly contributing to your research projects and publications. It is thus essential to balance those activities. The purpose of this paper is twofold: i) illustrate some of the benefits of being involved and, most importantly, discuss how to get there; and ii) give some concrete advice and rules to keep this involvement as effective and controlled as possible in order to serve the community and receive benefits in return without hampering your research activity.

In scientific societies or associations, many tasks are accomplished by individuals who volunteer their time. Even tasks that appear to be merely administrative or clerical are essential for the scientific community and will make a difference in your field. In those volunteer organizations, projects are often driven by a single person or a very small team. Consequently, volunteers often have to take initiative and take things into their own hands. That is the context in which these rules should be of particular interest.

I have been involved in the Student Council of the International Society for Computational Biology for five years, progressively taking on more responsibilities, in particular in the organization of conferences (co-chair of the symposium in Boston in 2010 and chair of the first European symposium in Ghent in 2010), but also more generally in the Student Council (I was secretary—one of the elected leaders—of the Student Council in 2009). In addition, I created the French Regional Student Group (RSG-France), which I chaired for two years. This paper is based on my experience in the bioinformatics community, but also on associative involvement I had outside science. Most examples are taken from the bioinformatics community, but I believe the rules are rather general and apply for other communities.

## Rule 1: Collect Information

Maybe you are not sure whether you want to get involved or not and which kind of involvement is possible and would be interesting for you. The first thing to do is certainly to ask people around you about their experience in various associations and committees, should it be in your scientific community or other communities. You can ask them about the kind of involvement they have or had and what they like or dislike about it. Which were the benefits? Which were the problems? Would they do it again? All these questions can help you get a more concrete idea. In addition, you can search on the Internet and look for information about societies or associations you are interested in, if they exist. If they don't, it can also be good to create something new, but that is more challenging and may not be appropriate for a first experience.

## Rule 2: Define What You Want and Expect

It is important to know why you are getting involved and to define a clear goal. This will help you keep the motivation. For instance, you want to be part of a team of international students to improve your communication skills, or you want to learn how you can raise funds and contact sponsors. Maybe you want to get experience in organizing a conference or simply meet new colleagues all around the world. Defining what you will get or expect to get from the involvement is certainly a good idea. You might realize afterwards that you actually got very different benefits from what you were expecting, but it is good to think about it at first.

## Rule 3: Define Your Boundaries

To keep the balance between your activities you need to define clear boundaries, in particular to what extent you want to get involved. If you don't know what you are doing, you don't know when to stop. This is true for the daily work when you are wasting a lot of time simply because the task is not clear. But it is also valid for the duration of your involvement. It may be a good idea to decide beforehand when you want to stop. Do you plan to be involved two years? Three years?

**Citation:** Michaut M (2011) Ten Simple Rules for Getting Involved in Your Scientific Community. PLoS Comput Biol 7(10): e1002232. doi:10.1371/journal.pcbi.1002232

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** October 27, 2011



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: magali.michaut@iscb.org

Until you get your PhD? Until you finish your postdoc or any other project? It may be easier to get involved after you have settled in your current place and project, as opposed to during phases of transition.

## Rule 4: Jump into the Pool and Get Involved

Now you want to get involved, you know why, and you have a goal and boundaries. But how can you actually start? Keep in mind that it may be enough to simply be open to any good opportunity that can unexpectedly happen. Haven't you been asked already to help out with the organization of an event or the reviewing of some abstracts? Otherwise, you will need to be proactive to get involved, and there are many ways to start. For instance, you can send an e-mail to a committee chair or the society chair asking questions about how it works or how you could help. You can even indicate your interest if you have some ideas or know what you would like to do, but it is certainly not required. Don't hesitate to contact people and just ask if there is anything you can do to help. Help is often needed and very appreciated. You can also attend the annual meeting of the society, join a committee, or participate in mailing list discussions. Even when you are already involved you can be proactive about taking on more responsibilities. If would like to do more, or change what you are working on, let people know and offer to do something different or new. It is always very motivating for the team to see that volunteers want more responsibilities.

## Rule 5: Let Other People Know What You Want to Do

Everybody has different interests and it is key to know them to build a team as effective as possible. If Joe hates contacting potential sponsors but likes writing meeting reports, he will be happy to know that William would rather be part of the fundraising effort and hates writing reports. Thus, be clear about your interests for the benefit of everybody. Following this idea, it is important to be clear with yourself and with others about what you can or can't do. You have to realize that you are part of a team. The point is not to do everything, or to take as many tasks or responsibilities as possible to show you are very much involved. The point is to commit to what you can do and to do it (and do it as well as you can). If you have some more time, you can always ask for more, help on other tasks, and get more involved. But if you can't deliver what you signed up for, you penalize the team and the work of other people. You can think of it as a soccer team—if you commit for a game and don't show up, the team is stuck.

## Rule 6: Dedicate Regular Time

It is extremely important to work regularly even when you are busy. It is indeed very likely that your research will take up all the time that is not firmly reserved for other activities. Thus, if you don't take your involvement as seriously as your research, you will never get anything done. When you feel overwhelmed, postponing everything for later when you expect to have more time is generally not a good strategy, because you will always be busy. It is often the case that 10 or 15 minutes on a project can be enough to get the next step done. Think about where you are and what is the next step. Maybe you just need to send an e-mail to ask about the quotes Jack had to get, or remind this keynote speaker about the picture he has to send. However, we still have some periods when it is more difficult than usual to dedicate the smallest amount of time. In that case, be clear about it and try to give your expected schedule and deadlines in advance so that other people on the team can adjust.

## Rule 7: Organize Your Time

Since you can't spend all your time on your community involvement and want to maintain a balance with the activities directly related to your research projects, it is essential to get organized. You can decide in advance how much time you want to dedicate and track the time you actually spend on your various activities. You might realize that some tasks take much more time than you were expecting or, conversely, are much faster to perform than you initially thought. The more you do it, the more accurate you become in your time estimates. This will enable you to know precisely which responsibilities and tasks you are able to handle and to be reliable in your commitments. As part of your schedule, you also want to define realistic milestones and deadlines, and stick to them.

## Rule 8: Work in a Team

Unless you are really working on a project alone, you will likely be part of a team and you should take advantage of it. Thus, don't take all the work for you, and remember that you are not alone. Keep in mind, particularly if you lead a team, that you need to distribute the work, delegate some tasks to others, and ask for help when you need it. In general it is good to assign a single responsible person and a deadline for each task. Working with other people is also an interesting way to get feedback on your work and ideas. Even though it usually takes more time, it is a good idea to suggest a discussion and take the opportunity to get comments on your ideas, actions, and concerns. That is what teamwork is about. Finally, this is probably more geared towards leaders, but it is extremely important to be able to get the best out of a group of different and complementary volunteers. Identify the strengths and weaknesses of your team workers and help everybody achieve their best based on their interests and skills. Identify and respect the differences of the people in the team. In particular, in international associations you will likely be interacting with people from all over the world who may have cultural differences in work styles, expectations, and ways to communicate. In line with this, it may be useful to provide an action item list with concrete tasks that allows people to find where they can help in the project.

## Rule 9: Encourage Others to Get Involved

Don't hesitate to let your colleagues know about your involvement. The point is not to show them how great you are doing and that they should do the same. But it is very likely that many people are not aware of this kind of involvement and don't realize how useful it is for the community and for you. Explain the work you are doing and what you get from it. You can encourage your colleagues to play an active role in the scientific community. If you think that someone would be effective in some specific task, tell him or her so. Sometimes people don't realize that they are good in specific tasks that seem complicated for others. For instance, you can ask Averell, who has very good graphical skills, to work on the design of various documents, flyers, or posters. Since the organization is composed of volunteers, it is often the case that people have to step down from their position when their job situation changes. Thus, it is important to have other volunteers who can take over. But it is also important to get new people to bring fresh ideas, new perspectives, and different ways to work. When you start to know people and have experience working with them, for example, in organizing a conference, you can be very effective doing similar tasks again.

Nevertheless, it is rewarding to get new people involved and to have new comments from outside, even if it seems more complicated and takes more time. Last but not least, you should guide interested people to get involved. Many people would be happy to help but don't take the time to actually start, or don't feel confident enough. If you mentor them in the beginning, it might be enough for them to get into it.

## Rule 10: Enjoy as Much as Possible

What you like, you will do great without specific effort. If you know why you are doing it and if you enjoy it, you will take the time to do it, and you will do it well. And if you don't like it anymore or get bored, then finish your commitments and discontinue that activity. Of course, I should emphasize here that you have to finish your commitments first (see teamwork comments above)!

I hope I managed to illustrate that getting involved in your scientific community is not only extremely rewarding for you, but also possible for everybody, and that simple rules can help you balance your activities. There is a lot to do, various tasks for various people and at different levels of involvement. Every experience is of course different, and I would be glad to hear about your experience, should it be similar or very different. It is possible that you will have a bad experience or that something you try will not work out. In that case, don't be discouraged and try something else. Your experience can also simply be different from what you were expecting, but in the end, it is always a good experience. After all, experience is what you get when you didn't get what you wanted.


## Acknowledgments

I would like to thank Nils Gehlenborg for detailed and insightful comments on this paper.

Editorial

# Ten Simple Rules for Getting Help from Online Scientific Communities

**Giovanni M. Dall'Olio[1]*, Jacopo Marino[2], Michael Schubert[3], Kevin L. Keys[1], Melanie I. Stefan[4], Colin S. Gillespie[5], Pierre Poulain[6,7,8], Khader Shameer[9,10], Robert Sugar[3], Brandon M. Invergo[1], Lars J. Jensen[11], Jaume Bertranpetit[1], Hafid Laayouni[1]**

**1** Institute of Evolutionary Biology (UPF-CSIC), Departament de Ciències Experimentals i de la Salut, Barcelona, Spain, **2** Institute of Organic Chemistry Universität Zurich, Zurich, Switzerland, **3** EMBL-EBI, Wellcome-Trust Genome Campus, Hinxton, United Kingdom, **4** California Institute of Technology, Biology Division, Pasadena, California, United States of America, **5** School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne, United Kingdom, **6** DSIMB, INSERM, U665, Paris, France, **7** Univ Paris Diderot, Sorbonne Paris Cité, UMR-S665, Paris, France, **8** Institut National de la Transfusion Sanguine, Paris, France, **9** Division of Cardiovascular Diseases, Mayo Clinic, Rochester, Minnesota, United States of America, **10** National Centre for Biological Sciences (TIFR), GKVK Campus, Bangalore, India, **11** NNF Center for Protein Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark

## Introduction

The increasing complexity of research requires scientists to work at the intersection of multiple fields and to face problems for which their formal education has not prepared them. For example, biologists with no or little background in programming are now often using complex scripts to handle the results from their experiments; vice versa, programmers wishing to enter the world of bioinformatics must know about biochemistry, genetics, and other fields.

In this context, communication tools such as mailing lists, web forums, and online communities acquire increasing importance. These tools permit scientists to quickly contact people skilled in a specialized field. A question posed properly to the right online scientific community can help in solving difficult problems, often faster than screening literature or writing to publication authors. The growth of active online scientific communities, such as those listed in Table S1, demonstrates how these tools are becoming an important source of support for an increasing number of researchers.

Nevertheless, making proper use of these resources is not easy. Adhering to the social norms of World Wide Web communication—loosely termed "netiquette"—is both important and non-trivial.

In this article, we take inspiration from our experience on Internet-shared scientific knowledge, and from similar documents such as "Asking the Questions the Smart Way" [1] and "Getting Answers" [2], to provide guidelines and suggestions on how to use online communities to solve scientific problems.

## Rule 1. Do Not Be Afraid to Ask a Question

Some people are afraid of asking a question in public, for fear of appearing ignorant or foolish. Other people worry about their ability to express the question proficiently or with the correct grammar.

Actually, asking a question in a public website is a good thing. First, the process of composing a message to explain a problem is itself a great exercise. Second, it is a great way to learn faster, and to enter into contact with people from different fields. Third, and more importantly, your career will be difficult if you do not learn how to get help from other people.

As Albert Einstein once said, "The important thing is not to stop questioning. Curiosity has its own reason for existing" [3]. Asking the right questions should always be a priority in science, and online communities are a good place to practice.

## Rule 2. State the Question Clearly

The key to getting a good answer is to ask the question in a clear and concise way. If your question is too long, many people simply will not read it. On the contrary, if your question is too short, people may interpret it incorrectly and give you an erroneous answer.

A way to keep your questions short and concise is to systematically break down the problem into smaller parts. This can help you to decide where to seek help, and how much to seek. If you feel your problem is composed of multiple questions, then post as many messages as needed. You should start a separate discussion thread for each of the problems you want to solve, avoiding mixing messages about different topics together.

On the other hand, you should provide enough details so that people can answer you without having to ask you for additional explanations. Read the message you wrote carefully, and think about which details you forgot to include. A reader should be able to answer you just by reading your initial message, without having to look at the rest of the discussion, or at what other people already have said in response.

Some examples of non-concise questions and how to improve them are shown in Text S1. Spend as much time as you need in preparing your initial message: this will save time later and will lead you to find the best solution more easily. Many people are surprised to see how some-

**Citation:** Dall'Olio GM, Marino J, Schubert M, Keys KL, Stefan MI, et al. (2011) Ten Simple Rules for Getting Help from Online Scientific Communities. PLoS Comput Biol 7(9): e1002202. doi:10.1371/journal.pcbi.1002202

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** September 29, 2011



**Funding:** GMD is supported by grants SAF-2007-63171 and BFU2010-19443 (subprogram BMC) awarded by Ministerio de Educación y Ciencia (Spain), the Direcció General de Recerca, Generalitat de Catalunya (Grup de Recerca Consolidat 2009 SGR 1101) to JB. KLK is supported by a 2010–2011 Fulbright student research grant to Spain. LJJ is supported by a donation from the Novo Nordisk Foundation. JM is supported by funding from the University of Zurich. MIS is supported by a long-term fellowship from EMBO. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: giovanni.dallolio@upf.edu

times, in thinking about how to pose the problem, the answer reveals itself!

## Rule 3. New to a Mailing List? Learn the Established Customs before Posting

A common error is to rush into a web forum and start asking something without understanding how its web interface works and which people use the resource. Instead, a good habit is to spend a few days, after having created an account, reading the discussions published and practicing with the web interface. You will see which people use the forum or mailing list, which rules of netiquette are used, which kind of questions are asked, and how much time it takes to obtain an answer. For this reason, it is a good idea to subscribe to a few mailing lists or forums on your topics of interest even when you do not urgently require anything from them. This will show you the concrete ways in which people post messages.

Remember that you may have to use a different language depending on the audience you are addressing. For example, some technical terms may be understood in one mailing list or community but not in others. People who do not study genomics might not immediately know how to respond to questions about GWASs, SNPs, or STRs (genome-wide association studies, single nucleotide polymorphisms, and single tandem repeats, respectively).

## Rule 4. Do Not Ask What Has Already Been Answered

People in general do not like to repeat their explanations. Before posting a question, use a search engine to see if a similar question has been asked previously. You should post a new question only if the answers you have found are not satisfactory. In case you decide to post a new question, cite the previous answers and explain why they are not sufficient to solve your problem. This demonstrates that you have already researched the answer on your own. Most discussion forums or mailing lists also have a searchable archive, which should be consulted before posting a question.

## Rule 5. Always Use a Good Title

People like to quickly skim through titles, looking for questions within their expertise that they are able to answer. So, you will have to be good at catching the attention of the readers that can help you. Use a clear and concise title, so that readers can decide whether they are able to respond to your message without having to read the whole message.

An approach to choosing a good title is to think of a hypothetical web search query that you would use to find a solution to your problem. For example, where you might search for "format BLAST database," an adequate title for a forum post could be "How do I format a BLAST database?" or "Formatting a BLAST database." More specificity, within reason, is preferable.

At the same time, it is important not to waste the time of the people who are not able to help you, and are not interested in what you are writing. Refrain from attempts to attract attention with titles such as "Help me" or "Urgent." People usually do not appreciate these kinds of titles because each forum member must then view the post in order to understand what you are asking. If you use incorrect titles, your message may be censured or closed by the moderators, and you may be forbidden to use the resource.

Some examples of good and bad titles are shown in Text S1.

## Rule 6. Do Your Homework before Posting

People in an online community are willing to help, but are not there to work for you. You should always show that you have first tried to solve your problem by yourself. Explain clearly what you have done, and describe the approach that you took.

When asking for help to solve an assignment, always explain how you have tried to solve it. Many students from bachelor programs use web forums and mailing lists to copy-paste the assignments given by their teachers, and call on other people to show them how to solve them. This behavior is not well received and can bring you a bad reputation.

However, you can nonetheless ask for help on how to solve an exercise if you demonstrate that you have made some effort in solving it. Show what you have done so far, and why you think it is not correct. Ask other people to check your solution, not to give the solution to you.

When asking about a programming issue, do not expect other people to write a whole program for you: rather, post an example of the code that you have written and where you are stuck. Include an example of the input and the expected output of your program. If you receive error messages, also include the full output of the error. This will help the other users to inspect your logic, to test the code on their own computers, and to easily pinpoint the problem therein.

If you ask a question about a software package, make sure that the solution is not already answered in the user manual or the Frequently Asked Questions (FAQs) before bringing your question to a forum. Also, declare that you have already checked these sources.

If you really need another person to write a program or a task for you, then explain that you are looking for a collaboration, and say how you will acknowledge a correct answer. If you explain everything well, your reputation online will also improve.

## Rule 7. Proofread your Post and Write in Correct English

Using correct grammar is important. Readers will be more likely to answer if the question is clear and correctly posed. Your grammar does not need to be academic, but it must be intelligible to a broad audience. Avoid slang and abbreviations as much as possible, to show that you have made at least some effort in writing a clear message. Writing in capital letters or in unconventional styles, such as that of text messages, is usually unwelcome, and in the long term can deteriorate your reputation online.

Your message should be as concise as possible. You do not need to introduce yourself on every message; doing it only once will be enough. Be careful of using too many adverbs and adjectives, or unnecessary changes in verb tense, as they may make the text difficult to understand. Also, do not be afraid of repeating technical terms more than once, as using too many synonyms will only make the text more difficult to understand.

This rule may be the most difficult to follow for non-native English speakers. A good approach is to spend some time reading the messages written by other users of the forum or the mailing list and follow their example. Search for a question similar to what you want to ask, and use it as a model; you may even copy and paste some portions of the text if it helps you to formulate a correct question.

## Rule 8. Be Courteous to Other Forum Members

Members of a discussion forum are usually unpaid volunteers who offer their time and expertise by volition and not by obligation. They are therefore not obliged to answer any questions at all.

Maintaining civil and polite conversations fosters an environment that encourages people to contribute. You must remember that forums are as human as their users, and you may sometimes receive a perfect answer written in an unfriendly tone. This can happen for various reasons: perhaps the same question was asked previously, or maybe the author was in a bad mood when writing. For your career, it is crucial that you not permit the discussion to degenerate into an argument. Even if you receive an impolite answer, stay calm and answer as gently as you can [4]. And remember the golden rule: treat other forum members as you wish to be treated.

One of the most impolite behaviors toward an online community is asking a question in multiple places at the same time. "Cross-posting", as this practice is called, can make two distinct online communities work through a solution for you when only one is needed; this is an abuse of forum members' time. If you have not received an answer and you believe that asking it in another place would get you one, provide a link back to the original discussion. Similarly, if you receive an answer in a different forum, report the answer to the original forum. Then, the people who helped you will know what the correct solution is and that you are no longer looking for it.

## Rule 9. Remember That the Archive of Your Discussion Can Be Useful to Other People

Messages in a mailing list or forum remain archived on the Internet. In certain situations, this can be a source of trouble: check the policy of your university or employer regarding posting on the Internet; avoid spreading embargoed information; and if possible, use your academic/corporate email address when registering, to keep your private life separated from your work.

Nevertheless, most of the time it is possible to make use of online communities without breaking any of your employer's rules. In these cases, the fact that an archive of the discussion remains publicly accessible is positive, as it becomes a useful resource for people searching for solutions to similar problems. Several knowledge archives are actively saving bioinformatics-related questions from open source projects. For example, questions about BioPerl [5] are kept in the GMANE (http://news.gmane.org/gmane.comp.lang.perl.bio.general) and Nabble archives (http://old.nabble.com/BioPerl-f13596.html).

Since an archive of the discussions remains available on Internet, it is good practice to conclude the discussion by indicating the correct solution to the problem exposed or by summarizing the suggestions received. If some of the answers that you received have proven to be wrong, do not be afraid of writing it in the online discussion: this will help other people avoid trying an erroneous solution. Even if you did not receive any useful answers, sacrifice a bit of your time to thank the people who tried to help you and to explain that you were not able to find a solution.

## Rule 10. Give Back to the Community

Have you found your answer? Great! As time progresses and you get more experienced in the respective field in which you asked your question, you might want to start contributing the knowledge that you have gained by helping people that are now in your previous position. Most online communities are very welcoming to new members, as they alleviate the work of more experienced ones. Also, as a new contributor, you might be able to see problems from a beginner's point of view. You do not have to contribute to the community by answering questions, as some communities have a "wiki-style" interface where you can contribute by editing, tagging, or flagging questions. In any case, following at least a few science-related mailing lists and contributing actively to them is a great way to come into contact with researchers working in your field, and over time can lead you to new collaborations and new opportunities for your career.

## Supporting Information

**Table S1** List of bioinformatics- and biology-related mailing lists and communities.
(DOC)

**Text S1** Examples of poorly posed questions, and how to improve them.
(DOC)

## Acknowledgments

This article has been written collaboratively as an open initiative based on the WikiGenes platform, at the address http://tinyurl.com/tenrules-onlinecommunities. We would like to thank Robert Hoffmann from the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, for making available the WikiGenes platform and for the support. We would also like to acknowledge Nelson Ndegwa from EMBL-EBI and Daniel Mietchen from University of Jena for contributions to the draft of the manuscript.

## References

1. Raymond ES, Moen R (2006) How to ask questions the smart way. Available: http://catb.org/esr/faqs/smart-questions.html. Accessed 31 May 2011.
2. Ash M (2006) Getting answers. Available: http://www.mikeash.com/getting_answers.html. Accessed 31 May 2011.
3. Miller W (2 May 1955) LIFE magazine.
4. Bourne PE, Barbour V (2011) Ten simple rules for building and maintaining a scientific reputation. PLoS Comput Biol 7: e1002108. doi:10.1371/journal.pcbi.1002108.
5. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz S, et al. (2002) The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12: 1611–1618.

Editorial

# Ten Simple Rules for a Successful Collaboration

**Quentin Vicens, Philip E. Bourne***

Scientific research has always been a collaborative undertaking, and this is particularly true today. For example, between 1981 and 2001, the average number of coauthors on a paper for the Proceedings of the National Academy of Sciences U S A rose from 3.9 to 8.4 [1]. Why the increase? Biology has always been considered the study of living systems; many of us now think of it as the study of complex systems. Understanding this complexity requires experts in many different domains. In short, these days success in being a biologist depends more on one's ability to collaborate than ever before. The Medical Research Centers in the United Kingdom figured this out long ago, and the new Janelia Farm research campus of the Howard Hughes Medical Institute in the United States has got the idea, as it strongly promotes intra- and inter-institutional collaborations [2].

Given that collaboration is crucial, how do you go about picking the right collaborators, and how can you best make the collaboration work? Here are ten simple rules based on our experience that we hope will help. Additional suggestions can be found in the references [3,4]. Above all, keep in mind that these rules are for both you and your collaborators. Always remember to treat your collaborators as you would want to be treated yourself—empathy is key.

## Rule 1: Do Not Be Lured into Just Any Collaboration

Learn to say no, even if it is to an attractive grant that would involve significant amounts of money and/or if it is a collaboration with someone more established and well-known. It is easier to say no at the beginning—the longer an ill-fated collaboration drags on, the harder it is to sever, and the worse it will be in the end. Enter a collaboration because of a shared passion for the science, not just because you think getting that grant or working with this person would look good on your curriculum vitae. Attending meetings is a perfect opportunity to interact with people who have shared interests [5]. Take time to consider all aspects of the potential collaboration. Ask yourself, will this collaboration really make a difference in my research? Does this grant constitute a valid motivation to seek out that collaboration? Do I have the expertise required to tackle the proposed tasks? What priority will this teamwork have for me? Will I be able to deliver on time? If the answer is no for even one of these questions, the collaboration could be ill-fated.

**Enter a collaboration because of a shared passion for the science . . .**

## Rule 2: Decide at the Beginning Who Will Work on What Tasks

Carefully establishing the purpose of the collaboration and delegating responsibilities is priceless. Often the collaboration will be defined by a grant. In that case, revisit the specific aims regularly and be sure the respective responsibilities are being met. Otherwise, consider writing a memo of understanding, or, if that is too formal, at least an e-mail about who is responsible for what. Given the delegation of tasks, discuss expectations for authorship early in the work. Having said that, leave room for evolution over the course of the collaboration. New ideas will arise. Have a mutual understanding up-front such that these ideas can be embraced as an extension of the original collaboration. Discuss adjustments to the timelines and the order of authors on the final published paper, accordingly. In any case, be comfortable with the anticipated credit you will get from the work. The history of science is littered with stories of unacknowledged contributions.

## Rule 3: Stick to Your Tasks

Scientific research is such that every answered question begs a number of new questions to be answered. Do not digress into these new questions without first discussing them with your collaborators. Do not change your initial plans without discussing the change with your collaborators. Thinking they will be pleased with your new approach or innovation is often misplaced and can lead to conflict.

## Rule 4: Be Open and Honest

Share data, protocols, materials, etc., and make papers accessible prior to publication. Remain available. A trusting relationship is important for the collaborative understanding of the problem being tackled and for the subsequent joint thinking throughout the evolution of the collaboration.

## Rule 5: Feel Respect, Get Respect

If you do not have respect for the scientific work of your collaborators, you should definitely not be collaborating. Respect here especially means playing by Rules 2–4. If you do not respect your collaborators, it will show. Likewise, if they don't respect you. Look for the signs. The signs will depend on the personality of your

**Citation:** Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3(3): e44. doi:10.1371/journal.pcbi.0030044



Quentin Vicens is a Howard Hughes Medical Institute Fellow at the University of Colorado, Boulder, Colorado, United States of America. Philip E. Bourne is the Editor-in-Chief of *PLoS Computational Biology*.

* To whom correspondence should be addressed. E-mail: bourne@sdsc.edu

collaborators and range from being aggressive to being passive–aggressive. For example, getting your tasks done in a timely manner should be your priority. There is nothing more frustrating for your collaborators than to have to throttle their progress while they are waiting for you to send them your data. Showing respect would be to inform your collaborator when you cannot make a previously agreed-upon deadline, so that other arrangements can be made.

## Rule 6: Communicate, Communicate, and Communicate

Consistent communication with your collaborators is the best way to make sure the partnership is going in the planned direction. Nothing new here, it is the same as for friendship and marriage. Communication is always better face-to-face if possible, for example by traveling to meet your collaborators, or by scheduling discussion related to your collaborations during conferences that the people involved will attend. Synchronous communication by telephone or video teleconferencing is preferred over asynchronous collaboration by e-mail (data could be exchanged by e-mail prior to a call so that everyone can refer to the data while talking).

## Rule 7: Protect Yourself from a Collaboration That Turns Sour

The excitement of a new collaboration can often quickly dissipate as the first hurdles to any new project appear. The direct consequence can be a progressive lack of interest and focus to get the job done. To avoid the subsequent frustrations and resentment that could even impact your work in general, give three chances to your collaborators to get back on track. After all, your collaborators could just be having a difficult time for reasons outside of their control and unanticipated at the time the collaboration started. After three chances, if it feels like the collaboration cannot be saved, move on. At that point try to minimize the role of your collaborators in your work: think carefully about the most basic help you need from them and get it while you can (e.g., when having a phone call or a meeting in person). You may still need to deal with the co-authorship, but hopefully for one paper only!

## Rule 8: Always Acknowledge and Cite Your Collaborators

This applies as soon as you mention preliminary results. Be clear on who undertook what aspect of the work being reported. Additionally, citing your collaborators can reveal your dynamism and your skills at developing prosperous professional relationships. This skill will be valued by your peers throughout your career.

## Rule 9: Seek Advice from Experienced Scientists

Even though you may not encounter severe difficulties that would result in the failure of the partnership, each collaboration will come with a particular set of challenges. To overcome these obstacles, interact with colleagues not involved in the work, such as your former advisors or professors in your department who have probably been through all kinds of collaborations. They will offer insightful advice that will help you move beyond the current crisis. Remember, however, that a crisis can occasionally lead to a breakthrough. Do not, therefore, give up on the collaboration too easily.

## Rule 10: If Your Collaboration Satisfies You, Keep It Going

Ever wondered why a pair of authors has published so many papers together? Well, it is like any good recipe: when you find one that works, you cook it again and again. Successful teamwork will tend to keep flourishing—the first paper will stimulate deeper and/or broader studies that will in turn lead to more papers. As you get to know your collaborators, you begin to understand work habits, strengths but also weaknesses, as well as respective areas of knowledge. Accepting these things and working together can make the work advance rapidly, but do not hurry: it takes time and effort from both sides to get to this point.

Collaborations often come unexpectedly, just like this one. One of us (PEB) as Editor-in-Chief was approached not just with the idea for these Ten Rules, but with a draft set of rules that needed only minor reworking. As you can see, we have obeyed Rule 8. ■


## Acknowledgments

We wish to thank Tom Cech for insightful discussions, and Chrysa Latrick, David Zappulla, Barbara Cohen, Emma Veitch, Catherine Nancarrow, and Hemai Parthasarathy for helpful suggestions on the manuscript.

**Author contributions.** QV and PEB wrote the paper.

**Funding.** The authors received no specific funding for this article.

**Competing interests.** The authors have declared that no competing interests exist.



**References**

1. Borner K, Maru JT, Goldstone RL (2004) The simultaneous evolution of author and paper networks. Proc Natl Acad Sci U S A 101: 5266–5273.
2. Rubin GM (2006) Janelia Farm: An experiment in scientific culture. Cell 125: 209–212.
3. Smalheiser NR, Perkins GA, Jones S (2005) Guidelines for negotiating scientific collaboration. PLoS Biol 3: e217.
4. Burroughs Wellcome Fund, Howard Hughes Medical Institute (2006) Making the right move. A practical guide to scientific management for postdocs and new faculty. Chevy Chase. Available: http://www.hhmi.org/labmanagement. Accessed 21 February 2007.
5. Aiken JW (2006) What's the value of conferences? Scientist 20: 54–56.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for successfully hosting an intern at a scientific software company

**Kristine Briedis, Yi-Shiou Chen, Scott Markel***

Dassault Systèmes BIOVIA, San Diego, California, United States of America

* scott.markel@3ds.com

Internships are increasingly popular choices for motivated students looking to enhance their skills, explore various employers, and get a head start on finding full-time employment. Although there are a few Ten Simple Rules articles from the intern's [1,2] and mentor's [3] perspectives, there are none written from the employer's. We'd like to share what we've learned over the years.

We work for a scientific software company, focusing on writing data pipelining software for our computational biology customers, primarily in pharmaceutical and biotech companies. Although we are informed by our specific experience, many of the following rules could be applied to internships in other fields, as well as academia. In general, we believe that these rules also apply to startup companies. The biggest difference one of us has experienced is that there may not be the luxury of having enough time to spend preparing tasks and working with interns. And all tasks may be on the critical path.



OPEN ACCESS

**Citation:** Briedis K, Chen Y-S, Markel S (2019) Ten simple rules for successfully hosting an intern at a scientific software company. PLoS Comput Biol 15 (6): e1007020. https://doi.org/10.1371/journal.pcbi.1007020

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** June 20, 2019



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Rule 1: Interview well

Start early. Most of the find-an-intern process is now taking place in the fall for summer internships. When seeking an intern, we must first recognize that, especially for undergraduates, we are unlikely to find exactly the background we desire. Most students will have either a computer science major or are studying biology. Some may be in bioinformatics programs, but in our experience, this is more likely an option for master's degree students. We also need to decide if we prefer undergraduates or graduate students. One aspect is the pay rate. Another is that graduate students are already starting to specialize. Clearly describe what the intern will work on and accomplish during the internship. Let them know whether any of the work will be publishable.

Where to find candidates? We attend career fairs and post openings online. We talk with previous interns, the schools and programs in which they studied, and colleagues and friends both in our company and in the community. Leverage your professional networks. If your academic contacts invite you to present to their classes or student groups, say yes.

This is in addition to the standard hiring and tracking pipeline in place at our company. Dassault Systèmes has a well-monitored approach to interns. Descriptions are posted online, and our human resources (HR) department attends career fairs nationwide. The company also maintains a database of past applicants that can be mined for promising candidates. We track by brand, e.g., BIOVIA, and geography. We are interested in knowing interns sought, interns hired, interns started, interns who successfully complete, interns hired by the groups where they interned, interns who were hired within the same brand but a different group, and interns hired by different brands. This information has been gathered for more than 15 years within the larger company. A small company, e.g., a startup, might not have the history or infrastructure to support this, although the same information is likely to exist informally.

The interview process may be stressful for the students. Many have not previously interviewed at companies. Keep in mind that undergraduates will probably not have the same skills one would look for in a new hire for a full-time position. Instead, they're still learning.

## Rule 2: Plan for arrival

Internships often have a very short duration, sometimes only 2 or 3 summer months. You and your intern have to be ready to start on the first day. They'll typically be looking to you, as this is often very new to them.

Know what you want them to learn or become familiar with during those first days and weeks. Choose projects in advance (see Rule 5). Think about having your intern sit near team mentors. It makes answering questions more convenient for both of you. Talk with others who have had interns, especially at your own company. See what suggestions they have, including what didn't work so well. You may also need to provide the usual company overview for your intern; HR may not provide the same kind of new-hire orientation, or any at all, for interns.

## Rule 3: Introduce environment and tools

Many interns have not developed software in a commercial environment. They may not be familiar with common tools for bug and feature tracking, source code control, continuous build/integration, etc. Providing short tutorials and online help will enable your intern to be productive faster. They'll learn best by doing. You might want to provide a sandbox environment or make sure interns are appropriately monitored, especially when it comes to code check-ins.

Another aspect of a commercial environment is the commercial part. This may be new to the interns, as they may not have interacted much with private industry. They may need to sign nondisclosure agreements (NDAs), which is almost always a first for them. Another point of caution is customer meetings/calls. The shift from a more open academic environment to a commercial one often entails being a bit more circumspect in conversations. Interns typically need to be coached on this if you wish to include them in customer meetings. Alternatively, you might decide it is best to leave them out of confidential conversations.

## Rule 4: Mentor commercial software development skills

Provide a guided tour of your organization's best practices, including agile methods, if applicable. Our company has a good training checklist for newly hired software engineers provided by the quality management system team that we also use for interns. If possible, having an intern start at the same time as a new permanent employee allows for some reduction in needing to cover introductory material about the company, processes, tools, and tasks. It is critical for them to develop a good working knowledge of your source code control and issue tracking systems. Introduce task planning and code reviews. Pair programming and debugging will naturally go together. Interns can learn a lot by watching team members work through a problem, e.g., use of design patterns, test cases, and iterative development. We have our interns develop new functionality in a series of iterative steps, adding complexity as they show they've mastered the simpler parts of the problems.

## Rule 5: Choose tasks wisely

Start with small tasks to provide a sense of accomplishment. You'll probably also want to start with noncritical tasks so that the rest of the team isn't blocked. Use bug fixing as an opportunity for the intern to become familiar with existing functionality and design patterns. Fixing simple bugs can also provide a quick and instant reward for the intern. If possible, pick at least one task that will be in the upcoming release. We've seen faces just light up when they hear that something they've completed will soon be in the hands of paying customers; it's a sense of accomplishment for them way beyond working on homework problems and course projects. Have a mix of bug fixing, new development, and exploratory tasks. Is there a new technology that your team hasn't been able to spend time on yet? This can be a good opportunity to have an intern tackle that higher-risk project and provide added value to the company while being intellectually challenged. Allow for input from the interns; let them pick what they think they'd enjoy. We recently had a returning intern continue with a task from her previous summer. If the internship is of sufficient length, the intern can rotate through multiple teams or projects to get a broader view of the company's products.

Set clear expectations for what the intern is to accomplish during his or her stay. Because interns are employed for only a short time, we use a simplified version of our objective setting and performance evaluation process. We develop fewer objectives for interns and typically ignore the usual weightings. By interacting so closely with the interns, we can skip the intermediate performance evaluation meeting. Our exit process on the last day includes the intern's presentation, lunch, and a performance review.

## Rule 6: Help the intern learn about different parts of the organization

Many interns are surprised at how many other departments there are. We interact with, among others, quality assurance (QA), product managers, project managers, presales, sales, customer support, professional services, marketing, and information technology (IT). Learning about the other departments can improve an intern's understanding of the overall commercial software industry. We typically schedule a number of 30- or 60-minute meetings so that our interns can learn and ask questions. This also exposes the interns to different career paths. And it's an opportunity to meet people outside the development group. Encourage them to expand their network and connect on LinkedIn, Twitter, or other appropriate social media with the people they meet. Turning these meetings, in person or remote, into pizza lunches is a big hit.

## Rule 7: Interns should have fun

Make sure it's not all work and no fun. This can be hard to keep in mind if an internship coincides with the end of a release cycle. Organize social activities with other interns. HR departments can be a big help here. Off-site events are always a favorite. This past summer, our interns spent an afternoon at a local escape room. Be sure to document (photograph/video) such events for the end-of-summer presentation (see Rule 9) or future intern marketing. Encourage team members to look for opportunities to involve interns. Ping pong and pool tables provide good break options. A warning sign is noticing that your intern is always eating lunch alone. Worse is having someone else point it out. Arrange lunches for your intern with other colleagues if need be.

## Rule 8: An intern is not a temp contractor

Make sure the intern leaves having learned something. Spend quality time with the intern; don't assign tasks and then disappear. Provide opportunities to ask questions. Elicit questions. Students often don't know what they don't know. For them, problems that are hard, impossible, or unknown can all blur together. It can be hard to discern the difference and know how to start or continue. Ask leading questions to help an intern work through a problem without giving the answer. Remember that your organization's, and your team's, reputation will be shared with classmates when the intern returns to school. This can impact future hiring opportunities with that school.

## Rule 9: Host a "What I Did This Summer" presentation

It's a great way to summarize what's been learned and accomplished. The presentation can be used by the intern in classes or future interviews. We also use them to show new interns what their predecessors have done. Invite people outside of the immediate development team, e.g., research and development (R&D) leadership, HR, and people the interns have met or worked with (see Rule 6). These presentations are an excellent opportunity to communicate success to the rest of the organization.

## Rule 10: Get and give feedback

"So, how did we do? What can the company do better next summer?" Evaluate the intern as a possible future hire or a returning intern. Do you want them back? Do they want to come back? This is presumably one of the reasons your company has an internship program. Keep in touch with the intern. If the internship was a positive experience, offer to provide a letter of recommendation for future job and scholarship applications. Ask for intern candidates and new hire referrals. Find out when they're back in town during school breaks. Most won't say no to a free lunch with the team!

## References


1. Aicher TP, Barabási DL, Harris BD, Nadig A, Williams KL (2017) Ten simple rules for getting the most out of a summer laboratory internship. PLoS Comput Biol [cited 2018 Nov 19]; 13(8): e1005606. Available from: 10.1371/journal.pcbi.1005606 PMID: 28817622
2. Chen B (2014) Ten Simple Rules for Internship in a Pharmaceutical Company. PLoS Comput Biol [cited 2018 Nov 19]; 10(5): e1003600. Available from: https://doi.org/10.1371/journal.pcbi.1003600. PMID: 24811706
3. Masters KS, Kreeger PK (2017) Ten simple rules for developing a mentor–mentee expectations document. PLoS Comput Biol [cited 2018 Nov 19]; 13(9): e1005709. Available from: https://doi.org/10.1371/journal.pcbi.1005709. PMID: 28934208

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for providing a meaningful research experience to high school students

**Emily A. Lescak[1☯], Kate M. O'Neill[2,3☯]*, Giovanna M. Collu[4], Subhamoy Das[5]**

**1** Department of Biological Sciences, University of Alaska Anchorage, Anchorage, Alaska, United States of America, **2** Institute for Physical Science and Technology, University of Maryland, College Park, Maryland, United States of America, **3** National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, United States of America, **4** Department of Cell, Developmental and Regenerative Biology and Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **5** Department of Neurosurgery, Stanford School of Medicine, Stanford, California, United States of America

☯ These authors contributed equally to this work.
* oneill.katem@gmail.com





**Citation:** Lescak EA, O'Neill KM, Collu GM, Das S (2019) Ten simple rules for providing a meaningful research experience to high school students. PLoS Comput Biol 15(4): e1006920. https://doi.org/10.1371/journal.pcbi.1006920

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** April 25, 2019



**Funding:** The authors received no specific funding for this work, but they would like to acknowledge the funding that supported them while writing this article. EAL was funded by NSF PRFP 1611913. KMO was funded in part by MURI grant #FA9550-16-1-0052 to Dr. Wolfgang Losert at UMD and in part by the Intramural Research Program at NIH, NINDS, via grant Z01-NS003013 to Dr. Edward Giniger. SD was supported by the 2018 ChEM-H Postdoc Seed Grant and startup funds from the Kaltschmidt lab. The publication of this work was supported in part by the UMD Libraries' Open Access Publishing Fund and in part by the aforementioned MURI grant. The funders had no role in study design, data collection and analysis,

## Introduction

Much has been written about designing research experiences for undergraduate students [1–4], but what about providing meaningful experiences to high school students? There are many formal opportunities for high school students to conduct research, but early-career scientists and principal investigators (PIs) do not necessarily have much experience working with this age group, which presents different opportunities and challenges than working with undergraduates. Thus, we present guidance in this Ten Simple Rules article on how to be an effective research mentor for high school students based on our experiences as early-career biologists and our formal mentor training.

Studies show that students—and the general public as a whole—have a narrow view of what a scientist is, does, and looks like [5, 6]. The opportunity to work in a research group may be the first time that high school students encounter a "real scientist." Likely, it is also their first chance to peek inside the black box that is scientific research—something they may only know from the media. They will experience firsthand what it is like to work in a research environment (whether they are doing experiments or computational work) and will likely be surprised by how communication and collaboration not only are necessary to the scientific process but also make research more rewarding. Performing scientific research gives students the opportunity to witness the practical applications of concepts they have been taught in school and to observe how the experimental and analytical work done in research settings builds upon what they have learned in the classroom. Importantly, they will also experience the excitement and challenges of investigating open-ended questions without predetermined answers. Authentic research experiences can empower students to pursue research opportunities as undergraduates and to consider careers in science, technology, engineering, and mathematics (STEM).

Engaging high school students in research and the process of doing science allows them to form meaningful relationships with mentors who can help them stay on track academically, serve as role models, and help prepare them for future careers. By working with high school students from the local community, mentors can bridge the gap between scientists and the general public and encourage students to attend their local university, which is a benefit for the mentor's institution, too. For high school students—particularly those who will be first-generation college students—getting comfortable on their local college campuses can make a meaningful impact on their educational goals. There are also opportunities for their

decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

supervisors, who are often early-career scientists (graduate students or postdoctoral fellows [postdocs]), to broaden their mentoring skills, improve their communication of the complexities of everyday science to a new audience, and learn how to develop tangible project goals that can be tackled within a finite period—all of which are excellent professional development opportunities.

Opportunities for high school students may be initiated either informally, through outreach with local schools, or formally, through an established program. We have compiled a list of programs, organized by state, that provide high school students with research experiences; please note that this list is not exhaustive. In general, placements range from the occasional half-day visit to year-long internships, and some placements are not necessarily local. Although the rules presented here are intended to guide mentors who will work with students for at least a few weeks, mentors working with students for shorter periods may also find some of these rules helpful.

Some universities and medical schools have volunteer offices or organized programs for bringing high school students into the laboratory, so check whether there are already connections to schools in your area through previous student placements. Moreover, when initiating contact with prospective mentees, consider the opportunity you have to make a meaningful impact in the lives of young people who come from historically underrepresented and underserved populations or underprivileged backgrounds. Scientific societies and funding agencies may have specific mechanisms for funding summer high school students, and many of these are intended for students who come from groups that are underrepresented in science. Example programs from the list above include the American Fisheries Society's Hutton Program and the Short-Term Experience for Underrepresented Persons at the National Institute of Diabetes and Digestive and Kidney Diseases. Some of these programs also provide stipends for the students, which relieves the additional pressure of needing to find a summer job. However you decide to bring a high school student into the laboratory, be sure to discuss with the prospective mentee what they hope to gain from the experience to make sure that your expectations are aligned before either of you commits to the placement.

It is important to recognize that working with high school students presents different challenges and opportunities than working with undergraduates. For example, high school students may be more enthusiastic than undergraduates about performing research because they have likely only engaged in simple lab exercises at school. However, they also have less scientific knowledge than undergraduates and likely are not able to spend as much time doing research because of schedule restrictions. These challenges can easily be mitigated by the mentor with some planning, and we have found mentoring high school students to be extremely rewarding.

If you decide to take on a high school student, we offer ten simple rules as guidance for providing the student with a positive experience while they are working with you. Although these rules were written with postdocs and advanced graduate students as the intended audience, we anticipate that they will also be helpful for PIs who have not yet hosted a high school student in their lab. In addition to these rules, we also recommend participating in mentor training through the National Research Mentoring Network or a similar program and familiarizing yourself with the literature on best practices in mentoring ([7–9] among many others) to strengthen your foundation in communicating and goal setting.

## Rule 1: Check with your institution's environmental health and safety/risk management offices to confirm the rules and regulations for working with minors

Anyone working in a research lab must be compliant with institutional safety regulations. It is important to be fully aware of the required paperwork, training, permissions, and other administrative steps before you reach out to potential mentees. Students performing computational work will likely still have some training to complete before they can begin working in the lab. As the primary mentor, it is your job to work with your PI and institutional officials (e.g., environmental health and safety officers, building managers) to find out what needs to be done for you to be able to work with the student and for the student to be able to work in the lab.

Any online training that the student can complete before they start in the lab will save valuable research time, but they will likely have to participate in on-site training too. You as the mentor may also be required to complete specific training for working with minors, and the student's guardian will likely need to sign consent forms. If feasible, schedule an initial, in-person meeting that includes the student, their guardian, the placement coordinator (if applicable), and the lab PI to explain the nature of the work and address any concerns. When the student does start working in the lab, and if they are doing experimental work, there might be protocols or procedures in which they cannot participate directly because of their status as a minor (e.g., working with vertebrates or using high-risk equipment), but you can involve them by allowing them to observe you during these tasks if it is safe and legal to do so. Make sure that you have personal protective equipment (lab coats, gloves) in the appropriate size. Finally, be sure to document any training you and the student complete and keep copies in your office and with the lab's personnel records.

## Rule 2: Make sure that you and your PI agree on reasonable, time-bound expectations and goals for the student's mentoring experience

Be proactive in planning the student's placement in the lab, and discuss with your PI how you plan to manage your time with the student. Have a conversation with your PI and other lab members (if appropriate) about both the concerns associated with taking on a high school student and how the mentorship can benefit the lab's research program, your professional development, and, of course, the prospective student! By taking these steps, you will ensure that the relationship between the student, PI, and other lab members is off to a good start and that you and your PI are on the same page regarding expectations for—and limitations of—the experience.

## Rule 3: Be realistic about your expectations for the student, and provide positive feedback

For the student to have a positive experience in the lab, it is important to set them up for success by designing a realistic project. You have to consider how much time they can actually commit to the project outside of school—and also how much time you have—and whether you will have a whole day with them or only smaller blocks of time. Student availability will vary depending on the kind of placement and its required time commitment. Discuss the student's school curriculum with them or (if possible) their teachers to ensure that the project is designed at an appropriate level. A good project will result in the student feeling that they have accomplished something, learned new information and skills, and contributed to the lab's larger goals by the end of their time with you.

An equally important aspect of your mentoring relationship is providing positive, constructive feedback. The student may not have confidence in their laboratory or analysis skills, because they will be new to research, so make sure to praise them for their work. Positive affirmations will help them gain confidence in their abilities, which is particularly important for women [10] and other groups underrepresented in STEM [11]. Inevitably, the student will make mistakes (we all do!), but make sure you highlight what they have done well. Then, together, brainstorm ways they can improve. These microaffirmations can go a long way in inspiring them to realize their own ability.

## Rule 4: Set goals early, and revisit them often

The student might have unrealistic expectations of what they can accomplish during their research experience because they are new to laboratory research. Thus, it is your duty as the mentor to explicitly set goals with both the best- and worst-case scenarios in mind and to manage expectations. We suggest you familiarize yourself with some of the resources available on goal setting [12, 13] to ensure that the goals you set for the student are realistic.

Set overall goals for the entire duration of the student's research experience as well as for smaller periods of time (e.g., weekly), and revisit them regularly. The overall goal could be as simple as learning a new technique or analysis method or as complex as answering a small scientific question. High school students are generally accustomed to structured approaches in their schools, so providing a structured plan will help them to be productive in the lab and not feel overwhelmed. It will be a learning experience for them to realize that experiments or analyses often present technical difficulties and that original hypotheses are not always supported. You can show them how you iteratively improve, how you learn lessons from difficult experiments or analyses, and how these factors influence your goals. You can use these instances as teaching opportunities and explain that troubleshooting and course correction are critical steps in the scientific discovery process. Finally, to put everything on paper, consider developing a document in collaboration with the student that outlines expectations for communication and goals for your time together [14]. In this document, make sure that you agree with the student and guardian how you will communicate and that the student understands that they need to let you know in advance if they are unable to be in the lab at their scheduled time. This will help them develop their understanding of professional norms and managing deadlines.

## Rule 5: Design a deliverable for the end of the experience

Just like high school students have finals or class projects at the end of the term, it is important for you to work with the student to produce a final deliverable at the conclusion of the research experience. Examples of deliverables include the following: (1) a short summary to be shared with their teachers or school newspaper; (2) a presentation to their science class; (3) a summary for their college application; or for longer-term placements, (4) a poster presentation at the university/institution or a local conference. Formal mentorship programs likely require a presentation in one of these formats. Establish this expectation at the start of your time together (see Rule 4), and set aside time for the student to start working on the deliverable as soon as possible. Provide accessible, relevant background literature so that they can begin learning on their first day in the lab. To ensure they stay on track, set checkpoints along the way so that the student can complete the deliverable on time. If they are not part of a formal program that has a planned presentation at the end, discuss which of the various options works best for them. The experience of summarizing and presenting their research—no matter the format—is not only a valuable learning opportunity in terms of understanding their own work but also important for developing communication skills. Bear in mind that the student will need guidance in best practices for presenting and synthesizing their work. Provide them with examples and resources to empower them to be successful, and start the process early to avoid unnecessary stress.

## Rule 6: Structure the student's time when they are in the lab

Do not assume that the student has experience with time management, because it is generally managed for them by their school. Spend time explaining how you plan your schedule and manage distractions. Encourage good practices, such as planning experiments or analyses in advance and filling in their lab notebook. For long-term placements (summer or year-long), be proactive about dividing your own time between working with the student and working on your independent research. Build in dedicated time for the student to read through protocols and any other information that you provide so that they can process and reinforce the knowledge they are acquiring. If the placement requires the student to spend several hours in a row working in the laboratory, pay attention to their energy levels, and be flexible about break times because they are likely accustomed to having breaks throughout the day at school.

To enhance the student's experience, also consider introducing them to other scientists and staff at your institution so that they can learn about different aspects of research and STEM careers. This can be transformative in college planning and can also expose them to career options they did not know existed. If you teach or engage in outreach or other aspects of service, consider allowing the student to attend your classes or meetings, if appropriate, so that they have a more complete idea of what a day in the life of a scientist is like.

## Rule 7: Help the student see both the forest and the trees

An important aspect of learning to think like a scientist is to understand the big picture (the forest) and how each experiment (the trees) meshes with those goals. Explain to the student the context of the project to which they are contributing, the big questions that you are trying to answer, and how their work fits into the lab's overall goals. It can be easy for students to undervalue the work they do, particularly because day-to-day lab work tends to be iterative with incremental gains. Impress upon them the value of their work, and make sure they thoroughly understand each step. Consider also inviting the student to research group meetings so that they can better understand the broader picture of the work you are doing and the collaborative nature of research.

## Rule 8: Guide the student toward becoming independent in their work and taking ownership of their project

Performing scientific research in the laboratory requires a level of independence that is not as necessary in the classroom, and this may surprise the student working with you. To help them grow as a scientist, make sure to explain this difference at the beginning of your time together, and reinforce it often. Explain how, unlike projects designed for laboratory courses in school, there is no answer you "should" get in scientific research. There may be an answer that you are expecting—your hypothesis—but even the interpretation of those results can be open ended. Demonstrate to the student how you think outside the box when planning the next step or interpreting results, and encourage them to share their ideas. By brainstorming next steps together, you will teach the student by example how to take ownership of their project.

As their research progresses, hopefully the student is becoming proficient in experimental and/or analytical skills. Make sure that you are available when they are doing experiments and analyses, and be sure to guide them fully through a technique the first few times by showing them first and then doing it together until they feel comfortable. At this point, you will still need to supervise them to make sure they are working safely in the lab or setting up their analyses correctly. Build in reflective checkpoints so that the student can track their progress. They will likely have many questions at first and may not understand the purpose of each step—

science is not always intuitive. As the student becomes more independent, ask them in an informal and nonintimidating way why they do a certain step in an experiment or analysis. When they do make a mistake, address it right away, and assure them that it is part of the learning process. Teach them how to document any errors and resulting mitigation through note-taking. You can impress upon the student the importance of taking detailed notes, but they will need guidance on how to keep a lab notebook [15]. You can help by checking their notebook regularly and providing feedback (see Rule 3). Explain how the documentation process is critical for reproducibility, and relate it back to the "lab reports" that they have done—or will do—in school.

## Rule 9: Show the student you are human

High school students may be intimidated by you or your lab-mates even though that is not the intent. To ensure that the student feels welcomed in the lab, make sure to introduce them to other lab members during lunch or during a regularly scheduled group meeting. To become more relatable to the student, have conversations with them about what you were like at their age, your hobbies and experiences at school, and how you got to where you are today. Sharing the challenges that you have overcome will help the student understand that they are not expected to be perfect. They are likely anxious about the possibility of making a mistake, ruining your experiment, or not making a good impression. Showing the student how you handle and learn from mistakes will take some of the pressure off them.

Also, allow the student to see that you have a life outside of your work. If you have a family or other caregiving responsibilities that you feel comfortable talking about, share them with the student. It can be transformative for students to see that scientists can manage personal and work responsibilities. Doing so will also humanize you and strengthen your ability to be a role model for a diverse range of students.

## Rule 10: Establish a long-term mentoring relationship

Finally, we suggest that you approach this mentoring experience as an open-ended one. High school students, regardless of the paths they pursue, will be embarking on profound transitions after graduation. Whether they know it or not, they could use a mentor throughout the process—for application review, general advice, and/or networking opportunities. If you have rapport, there is no reason that your mentoring relationship must end when the student stops working in the lab. However, it is possible that the student may be too shy or feel bad about asking for more of your time outside of the lab. Offer to keep in touch, and mean it. Make sure the student has a way to contact you. Follow up with their guardian, teacher, or placement coordinator 1 month or so after they have left the lab to see how they are doing. You never know the impact you can make. Good luck!

## Conclusion

Integrating a high school student into the lab has both challenges and benefits. It certainly takes time to explain concepts, teach techniques, and supervise their experiments and analyses. However, this investment has the potential to provide an invaluable opportunity for the student to engage in meaningful work and to open doors for their future educational and career opportunities.

If you are currently or soon will be a mentor to a high school student, reach out to Future PI Slack (twitter handle: @FuturePI_Slack). We have a #mentoring channel where we discuss best practices, provide advice on challenges, and share successes. We also encourage all scientists, regardless of career stage, to develop mentoring networks that they can rely on for advice, encouragement, and feedback.

## Acknowledgments

The authors would like to acknowledge Future PI Slack and the Genetics Society of America's Early Career Scientist Leadership Program for providing platforms for them to meet and discuss mentoring. EAL would like to acknowledge Amber Jaasma, Gary Snyder, and Elsa Hoppenworth for valuable feedback. KMO would like to thank ASP at UMD and Ellen Wu at Feinberg for their helpful feedback. GMC thanks the Center for Excellence in Youth Education at ISMMS for coordinating high school student placements.

## References

1. Aicher TP, Barabási DL, Harris BD, Nadig A, Williams KL. Ten simple rules for getting the most out of a summer laboratory internship. PLoS Comput Biol. 2017; 13(8):e1005606. https://doi.org/10.1371/journal.pcbi.1005606 PMID: 28817622
2. Cantelli G. Mentoring undergraduates: all you need to know to help your summer student. 2018 Jun 29 [cited 2018 Oct 17]. In: American Society for Cell Biology Post: Careers [Internet]. Available from: https://www.ascb.org/careers/mentoring-undergraduates-need-know-help-summer-student/.
3. Yu M, Kuo Y-M. Ten simple rules to make the most out of your undergraduate research career. PLoS Comput Biol. 2017; 13(5):e1005484. https://doi.org/10.1371/journal.pcbi.1005484 PMID: 28472033
4. Summers J. How to mentor undergraduates as a postgraduate, and why it's important. 2018 Jun 13 [cited 2018 Oct 17]. In: Naturejobs Blog [Internet]. Available from: http://blogs.nature.com/naturejobs/2018/06/13/how-to-mentor-undergrads-and-why-its-important/.
5. Nelson D. Public perception of scientists, and what we can do about it. ACS Comments. 2016; 94 (25):34.
6. Ruiz-Mallén I, Escalas MT. Scientists Seen by Children: A Case Study in Catalonia, Spain. Science Communication. 2012; 34(4):520–45.
7. Lee A, Dennis C, Campbell P. *Nature*'s guide for mentors. Nature. 2007; 447(7146):791–7. https://doi.org/10.1038/447791a PMID: 17568738
8. Woolston C. Why learning to mentor and teach is more important for US faculty members than publishing papers. 2018 Jun 1 [cited 2018 Oct 17]. In: Naturejobs Blog [Internet]. Available from: http://blogs.nature.com/naturejobs/2018/06/01/why-learning-to-mentor-and-teach-is-more-important-for-us-phd-students-than-publishing-papers/.
9. Great mentoring is key for the next generation of scientists. Nature. 2017; 552(7683):5.
10. Park LE, Kondrak CL, Ward DE, Streamer L. Positive Feedback From Male Authority Figures Boosts Women's Math Outcomes. Personality and Social Psychology Bulletin. 2018; 44(3):359–83. https://doi.org/10.1177/0146167217741312 PMID: 29157130
11. Chavous T, Leath S, Gámez R. Climate, Mentoring, and Persistence Among Underrepresented STEM Doctoral Students. 2018 Jun 25 [cited 2018 Oct 17]. In: Higher Education Today [Internet]. Available from: https://www.higheredtoday.org/2018/06/25/climate-mentoring-persistence-among-underrepresented-stem-doctoral-students/.
12. Doran G. There's a SMART way to write management's goals and objectives. Management Review. 1981; 70(11):35–6.
13. Conzemius A, O'Neill J. The Power of SMART Goals: Using Goals to Improve Student Learning. Bloomington (IN): Solution Tree Press; 2009.
14. Masters KS, Kreeger PK. Ten simple rules for developing a mentor–mentee expectations document. PLoS Comput Biol. 2017; 13(9):e1005709. https://doi.org/10.1371/journal.pcbi.1005709 PMID: 28934208
15. Schnell S. Ten Simple Rules for a Computational Biologist's Laboratory Notebook. PLoS Comput Biol. 2015; 11(9):e1004385. https://doi.org/10.1371/journal.pcbi.1004385 PMID: 26356732

EDITORIAL

# Ten simple rules for delivering live distance training in bioinformatics across the globe using webinars

**Denise Carvalho-Silva[1,2]*, Leyla Garcia[1], Sarah L. Morgan[1], Cath Brooksbank[1], Ian Dunham[1,2]**

**1** European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom, **2** Open Targets, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

* denise@ebi.ac.uk





**Citation:** Carvalho-Silva D, Garcia L, Morgan SL, Brooksbank C, Dunham I (2018) Ten simple rules for delivering live distance training in bioinformatics across the globe using webinars. PLoS Comput Biol 14(11): e1006419. https://doi.org/10.1371/journal.pcbi.1006419

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** November 15, 2018



**Funding:** Webinars at EMBL-EBI are funded by EMBL; we also gratefully acknowledge support for several webinar series for projects funded by the EU and UKRI: BioExcel webinars are funded by the European Union Horizon 2020 program under grant agreement 675728. CABANA webinars are funded by Global Challenges Research Fund - part of the UK AID budget (grant no. BBSRC-GCRF - VA: 75710). CORBEL webinars receive funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654248. RItrain webinars are funded by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 654156. Part of this work is funded by Open

## Introduction

Bioinformatics learning opportunities are now easily available face to face [1] or online [2]. As a rule of thumb, the former can (and will) trump the latter for its level of interactivity and engagement [3]. Most, if not all, students appreciate having the trainer (and classmates) available and close by. Their questions will get answered on the spot, on a case-by-case basis, with a personal touch. If the students happen to be in Europe (e.g., [4,5]) or North America (e.g., [6, 7]), they are in luck: there is no shortage of opportunities for such engaging encounters. Funding is often available for these students to attend face-to-face training. However, other parts of the globe tend to get neglected when it comes to live (and lively) face-to-face scientific training. Although capacity-strengthening initiatives, such as the Pan African Bioinformatics Network for Human Heredity and Health in Africa (H3ABioNet) Initiative [8], CABANA [9], Asia Pacific BioInformatics Network (APBioNet) [10], attempt to address this inequality, especially in low- and middle-income countries, scalability will always be an issue for face-to-face training. Online courses [11–13], however, allow training at scale, regardless of the trainees' location. Funding for travel is no longer a hurdle: the only requirement is access to a computer (perhaps a smart phone or tablet) and an internet connection. The course is taken in the comfort and convenience of the trainee's home, office, a library, or perhaps a coffee place with free Wi-Fi. However, on-demand access can be offset by lack of interactivity. Although online training portals often have chat rooms or other means of interacting with fellow learners or the course provider, discussions initiated this way often have a lag time.

Web-based seminars (webinars) offer the best of both worlds: they are run online and therefore at no (or little) cost for trainees, they can be scheduled at a convenient time for the target audience, the geographic distance between the trainer and the trainee is no longer an issue, and they allow for interaction between trainer and trainees at the moment of delivery. Webinars are short and straight to the point; the duration is usually no longer than 60 minutes. Questions are encouraged. Quick polls can be launched at any time for further interaction and getting to know the audience. Hands-on exercises can be provided, and follow-up webinars can be arranged for further discussions.

How can you achieve a stress-free and successful live streaming of bioinformatics training, which is interactive and available to everyone everywhere? Here are 10 simple rules that we have developed over the past five years of organising and delivering webinars [14]. Although our 10 simple rules are designed to deliver training on bioinformatics resources and projects, they can be easily applied to other domains. Due to the low cost, short duration, and flexible,

Targets. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

potentially global access, webinars can be used to train and/or promote a variety of themes in bioinformatics, computational biology, and computer science. Webinars will shorten the cycle time of your training and give you leeway to broaden your arsenal of content. You will be able to cover examples on protists, bacteria, plants—typically of great interest in low- and middle-income countries [15]—and have time to explore new trends in the application of machine learning, artificial intelligence, and blockchain in life sciences. Despite the possibilities of different contents, please be aware this article is not about selecting a training topic but rather on delivering training using webinars.

## Rule 1: Choose your webinar software wisely

There is no shortage of "best webinar software products" recommendations out there. You have options to suit every pocket, operating system (e.g., Apple, PC, Linux), and internet browser (e.g., Internet Explorer, Google Chrome, Firefox) and to accommodate small or big audiences. Free products have the disadvantage of limiting your audience size: there are a limited number of "seats" for attendees, typically 30 in a single webinar session. If you want to reach out to as many people as possible, free software is therefore unlikely to be an option. You should allocate some money in your budget to account for that. See Table 1 for our top five software programmes, with the minimum requirement of must-have features, other key specifications, and the URLs to learn more.

You must also take into account other perks of the software you purchase. These are some useful features of your platform-to-be:

- Can you track registration and attendance?
- Can you share the screen during the webinar?
- Can you launch polls in real time?
- Can you collect feedback after the session?
- Can you obtain detailed analytics?
- Can you text chat with the audience?
- Can you record the session?
- Can you assess the attention level of the audience?
- Can you follow up with emails?
- Can you share the webinar materials (slides, exercises)?

**Table 1. Top five webinar software programmes.**

| Webinar/training software | Software installation | Maximum capacity | Storage space | Operating systems compatibility |
|---|---|---|---|---|
| GoToTraining | Yes, local download | 200 | 2 GB | PC, Mac |
| GoToWebinar | Yes, local download | 1,000 | Unlimited | PC, Mac |
| Cisco WebEx training | No, directly through a web browser | 100 | 1 GB | PC, Mac, Linux |
| Zoom video webinars | No, directly through a web browser | 10,000[a] | 1 GB | PC, Mac, Linux |
| Myownconference | No, directly through a web browser | 2,000 | Unlimited | PC, Mac |

These five software programmes all share a minimum requirement of features, namely polls, evaluation, chat, screen sharing, recording, telephone integration, mobile device (with application), and custom registration (i.e., branding, logo).

[a]Maximum capacity can be unlimited if broadcast via Facebook Live or YouTube.

https://doi.org/10.1371/journal.pcbi.1006419.t001

- Can you integrate the software with other tools (calendars, social media channels)?
- Can you change the control of the screen in real time to a presenter other than yourself?
- Can you copy (or export) the results of evaluation surveys into other tools?
- Can you copy evaluation surveys from other tools into the webinar software?
- Can you easily get hold of the support team in the event of a "webinar catastrophe?"
- Can the participants dial in using a telephone line in the event of poor Wi-Fi connection?
- Can the participants join webinars on the go with a mobile device?

If you can answer "yes" to most or all of the questions above, consider purchasing a license for that software. If not many questions get "yes," think about which of the items above are a must for you. Which ones can you live without? Go out there, do your research, get a free trial, carry out a dry run (or several), and move on to our Rule 2.

## Rule 2: Pilot it with celebrities and friends

Now that you have the webinar software of your choice, you should pilot it to get familiar with the infrastructure. You will need a topic and an audience for this. The right topic is the one that you know a lot about (and/or work with) and there is a need for training on. The audience is a group of people who you know and who can give positive and constructive feedback.

At the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), we chose to pilot our webinar software with well-known and widely used bioinformatics resources (hence "celebrities"), namely PDBe [16] and Ensembl [17]. Moreover, both resources have had global user communities and well-established face-to-face training [18,19] and online presence [20,21]. Our friendly audience (hence "friends") were members of the Industry programme [22], have had a close relationship with our training activities, were generous with their time, and were keen on exploring alternative means of training.

If you are a computational biologist or bioinformatician, you can pilot your webinar software on topics such as next-generation sequencing data analysis and data visualisation. The audience of friends could be bioinformaticians or (computational) biologists you have worked with, or colleagues in the bioinformatics facility (institution) you work at. Because many research-performing institutes—from universities to companies—have experienced training and development teams, you can engage with them and bring them into the discussion from the beginning.

Make sure you run this pilot within the free trial period of the webinar software. Once the pilot is done, check how well your software performed against the list of perks from Rule 1. If it did not perform well enough for the license you have to pay, try a different software. If you need to run several pilots until you find the software that works for you, or if you are in shortage of friends who are willing to be your audience, you can pilot your webinar with just one person, e.g., a work colleague or facilitator—see Rule 6, or go solo altogether.

## Rule 3: Get a host on board

If face-to-face workshops are already part of your training portfolio, the chances are you have a list of people who have hosted your workshops in the past. Get in touch with the previous hosts and invite them to host a webinar, as they will help you to find the right audience (see Rule 4). Having a host on board will take the pressure off of you so you can focus on the important things such as crafting the syllabus, writing the abstract of the training, and creating the webinar registration link (from the webinar software of your choice—see Rule 1). Remind them of the advantages of hosting a webinar (e.g., save money, time, and effort). They just need to forward your webinar registration link to their contacts. Once you become more confident with online training and get the gist of it, you may want to ditch the need for a host. And jump straight on to Rule 4.

If you have never delivered face-to-face training before, you will not have a list of hosts. If that is the case, consider getting in touch with your network of work colleagues (past and/or present) and tell them about your training using webinars. They may be able to find a suitable audience for you. If you want to deliver your webinar to a much broader audience, share the registration link via mailing lists and social media (e.g., Twitter, LinkedIn, Facebook, Medium, blog).

## Rule 4: Find your audience

Before advertising your webinar, save a few spaces for members of your team (possible supporters at the Q&A session—see Rule 7), in case your webinar is a sell-out (i.e., all places get filled up). To cast your net as wide as possible and increase the number of registrants, advertise your webinar on social media [23], virtual conferences [24], and promotional events where online presence is encouraged such as festivals and open days, mailing lists (BioStars, Reddit), newsletters, and relevant journals. You can also go back to your friends (see Rule 2) and ask them to share the webinar registration link.

Keep checking the usernames, email domain, and number of registrants if possible. Watch out for likely bots by looking through their registration. If names and/or usernames contain random characters (alphabetical or numerical) only, it is likely to be fake registrants. You do not need to do anything with this at that moment. Carry on advertising the webinar widely, but do check the bot suspects on the day, especially if they do not turn up (see Rule 6). You might not be able to do anything against bots, but you will keep this in mind to get a more realistic expected number for genuine attendees.

## Rule 5: Prepare your content: Less is more

You love what you do—the bioinformatics resource or project you work on—so it is natural that you will try to cram in a lot of information you think is relevant. Do not! The audience will likely not remember every tiny detail you include in your webinar. And as a distance learning activity, it is less straightforward for assessing whether you have lost them or they are still with you with full attention (see Rule 6).

Start small by only giving an overview and some highlights of your topic so that you do not end up with a large number of slides to present and having to rush through them. You can offer follow-up webinars to cover additional content and train your audience on the finer details of your topic (this can be part of a series of webinars). Do not overcrowd your slides with detail; have them as visually compelling as you can. Consider adding live demos or play a recording (.mov, .mp4, .gif) showing some functionalities of your resource, either in a graphic or programmatic way.

It is good practice to build a couple of polls into your webinar to hold your audience's attention—these can be general questions like asking them which country they are listening in from, or they could be more specific to your topic, such as whether they have used the method, service, and/or resource that you are about to describe.

Include a slide on the logistics of the live webinar, such as:

- Webinar attendees will be muted.
- Materials (slides, exercises) will be available for download.
- Polls will be launched during the webinar.

- Attendees are encouraged to ask questions, which should be addressed to everyone.
- Questions will be answered at the end (via chat box or otherwise). Note: if you have a colleague with you, you could agree with them to answer any questions while you deliver the webinar. If that is the case, introduce anyone else who is assisting you at the session.
- The session will be recorded and shared.

At this stage, you also need to think about your post webinar survey. We will tackle this in more detail in Rule 9 but for now, just consider adding a "survey" holding slide at the end of your presentation; this will ensure that you do not forget to mention the survey at the end of your talk. Finally, practice your talk, make sure you will not run over, and check whether it flows smoothly throughout.

## Rule 6: Lights, camera, action

The big day has finally arrived. You may have a full house or just a handful of registrants. Both are great. Even if not many have registered, you will be recording your session (see Rule 8), so the video will be available for anyone to watch it. It is very unusual to get a full turnout—more typical is that 40% to 60% of registrants turn up. A few could be bots (see Rule 4 on basic tips to spot them), but most are likely to be people who were genuinely interested in attending but have failed to do so because something more pressing came up. Do not feel let down. Monitor the level of attendance in future webinars and try to establish whether there is a pattern. If you are suffering from frequent low attendance, consider the following: was the timing wrong? Did your webinar overlap with other events on the same topic? If you are webcasting to a single host, did it clash with a regular seminar or meeting taking place at the host organization? You could consider sending them a message saying "We were sorry you missed our live webinar; here is the URL link to the recording and materials. Do let us know if you have any further questions or if we can do anything to make our live webinars more convenient for you."

You should now get your laptop ready for the live session. Make sure you close down email clients, Slack, Skype chats, and so on. You do not want notifications popping up at the corner of your computer screen or beep sounds during your training session. Do not deliver your webinar from your regular desk. Book a room instead to ensure you are in a quiet place and will not be interrupted during the webinar. Have your laptop plugged in and use a wired internet connection. If neither is possible, make sure your laptop is fully charged and that the Wi-Fi internet connection is working. Remember some of the basics for ergonomics. Bring your keyboard and mouse, rather than using the laptop's unfriendly counterparts. Get the screen at eye level so that your head and neck are at a comfortable position. Having the right eye level means that attendees will be looking into your eyes rather than eyelids or forehead if you choose to have your built-in camera on. Feel free to experiment with this setting. There is no "one size fits all." You can have someone with you facilitating the webinar (e.g., opening and closing polls—see Rule 7—and answering the questions that get asked while you deliver the webinar). Having a facilitator especially when you first start delivering webinars will let you focus on the content rather than worrying on how to fix any possible technical issues that may arise, thus reducing the risks of things going awry. We recommend that the facilitator be in the same room as you. Once you have delivered more webinars and mastered some of the rules above, you can have facilitators remotely, or you can go completely solo, which is possible once you become more confident with webinars.

Some software has a beep sound to notify you when people join or leave the training session. Disable it. It is annoying for the attendees and distracting for you. Focus on delivering the session rather than on the sounds of attendees joining in later or leaving earlier. Let your audience know this is your first live, training webinar.

## Rule 7: Be engaging

A golden rule when it comes to training is to know your audience and engage with them. One of the ways you can do so during your webinar is to make use of polls—the webinar software you purchase should offer this functionality (see Rule 1). Polls are short and simple questions that you build before you go live. You can ask, for example, which species your audience works on, which computer language(s) they code in, whether they have used your resource, which other similar resources they have used, and so on. Since you are new to webinar delivery, it is best to practice these beforehand.

Once you launch the poll live, watch the votes coming in and the overall percentage of votes, wait a minute or so, and then close the poll. Thank the audience for answering the question and share the results with them. The answers will help you pitch the level of your webinar even if your slides are already uploaded and shared with the audience. If it turns out, for instance, that the audience is mainly of users who already know the resource, you can skip the basics. Or if you learn that none of registrants is computer savvy, you can skim through the programmatic access. Just make sure you and your presentation are flexible enough so you adjust the content to your audience on the fly.

In addition to regular polls, you can also throw in direct questions and tell the audience to respond via chat. Besides allowing you to be more spontaneous, direct questions can be used for cases in which answers do not fit in the simple multiple choice format of polls. If you want to know where your audience is listening in from, throw that in rather than using a poll because the latter typically does not allow more than four of five different options.

Another way to engage with your audience is encouraging them to ask questions, during or after the webinar, and via the chat box or the speaker. It is entirely your choice. Experiment with the different formats, then choose which ones you are more comfortable with. Our choice is to take all questions at the end. The reasons are 2-fold. Firstly, you may be about to cover what an attendee has just asked. Secondly, you have only typically 30 to 60 minutes to cover your content, and questions during the training may leave you short of time. If you have a colleague (or facilitator), you can introduce her/him and say that they will be happy to answer the question via the chat box while you talk. This is an exception because it is not always possible to have a facilitator. In the absence of one, check the chat box at the end and answer the questions once your presentation is over. Whatever your choice, make these housekeeping rules clear in your logistics slide (see Rule 5).

We also ask the audience to address the question to "everyone," preventing another attendee asking the same question. Besides, having the questions available to everyone allows for attendees to chip in (if they wish) or engage with the discussion, which adds an extra flavour to the engagement mix. Another advantage of having a written Q&A is that you can keep them as a record of the pain points and/or feedback.

If you choose to do the Q&A verbally, do so at the end of your webinar (for the same reasons outlined above) and unmute all the participants. Be aware of likely echoing once the microphones are no longer muted, especially if the participants happen to be in the same room. Also take into consideration that speaking may put some attendees off from asking questions, especially if the language of the webinar is different from their mother tongue. They will probably prefer writing the question down rather than speaking it out loud.

## Rule 8: Record the session and share it

Webinars are "for life, not just for Christmas"; i.e., they are not just for that single time of broadcast. Record your session to make it available to those registrants who did not make it to the live webinar. Share the recording more widely (perhaps by posting to a service such as

YouTube or Vimeo) and make it available to a much broader audience. Do not make the mistake of delivering the webinar and at the end realising you did not hit the "record" button. Press the button for both screen and voice even before the session starts. If you still forget to hit the record button, or if there are issues with the recording, there is nothing to stop you from delivering the same webinar again, this time with no audience, and making a clean recording available. Once the recording is done, go through it before sharing it. You may need to edit it (but do not spend too much time doing so) to remove long pauses, the start and the end of the recording, and the Q&A session. Once you have the final version, make the recording available from a website, e.g., the tutorial page of your own resource [25] or from a video-sharing website [26–29]. Video-sharing websites often have a social media element that can pull in a new audience previously unaware of your bioinformatics resource, so it is worth considering both approaches. Whichever approach you take, use your social media network to disseminate your new video. Regardless of the number of attendees in the live webinar, once the recording is shared, it will function as an online resource, which over time will reach a much wider audience.

## Rule 9: Get feedback and act on it

Hooray! You have delivered your webinar, the recording is online, and you are ready to move on to deliver more webinars. Before moving to the next, you need to assess how the first one went. Seek feedback on the style, the content, and/or the technical aspects of your training session. Evaluation surveys are the first channel of feedback that comes to mind. You can choose to send these as soon as the webinar is over, an hour later, or perhaps the day after. The sooner you send this, the higher the chances that the attendees will fill it out. It helps to know how to write surveys that work [30,31]. Our surveys are short (four to five questions) and have a combination of open questions and multiple choice (three to four options per question). We ask attendees to rate the session (from 1 [poor] to 5 [best]; be aware that some respondents may get confused and think 1 is best and 5 is poor). Precourse surveys are less likely to be representative of the audience and do not work for webinars because attendance rates are typically considerably lower than the number of registrants.

The Q&A at the end of the webinar (see Rule 7) is also a chance for the attendees to "voice" their final comments, which tend to be positive feedback. They can vary from "thanks," "enjoyed a lot this webinar," "I learned quite a lot" to "great presentation," "perfect introduction to this resource," "it looks like a really useful tool (amazing job)!" Make a note of them. Once your webinar is published (see Rule 8), monitor future engagement with it through comments and thumbs up or down on video-sharing websites.

Be objective when acting on the feedback, especially negative feedback. Some issues may be a matter of style (e.g., whether to show a thumbnail video window or not). Others can help you not to make the same mistake again or decrease its likelihood (e.g., once we launched a poll but forgot to close it, and the presentation froze). Look at the feedback as a standard normal curve: one may find it exceptional, another may find it mediocre; see what the majority says, and base your actions primarily on that. If you do not agree with the feedback, it may help to share it with colleagues for a second opinion. You can modify your materials (or delivery) accordingly. Besides feedback on content and style, participants may also highlight technical hiccups. Your software of choice may not be the best for their operating system or choice of internet browser. If that is the case, you can build recommendations into your registration confirmation email or housekeeping slides for future webinars. If you have had your camera on during the webinar, the feedback can be on the settings (e.g., light in the room being too bright or too dark, the wall behind the speaker too distracting, poor quality of the sound, and/or network speed). Take note of those issues and advise the forthcoming participants of some of these, so you (and they) do not run into similar issues. Overall, you will be able to act on the feedback received and adjust the content and level of detail if necessary in future webinars.

## Rule 10: It is not the end

Once webinars become part of your routine, you may tend to believe that this is it. You have nailed it. It is the end. Far from it. It is actually the beginning. The webinar can be a flavour of what is to come. Now that your audience has had a taste of the training that you can deliver, they may invite you to deliver face-to-face training. Or you might find that webinars are an effective means of providing updates on a bioinformatics methods or resources.

Things you may want to consider for future webinar training sessions are as follows:

- How often will repeats be required?
- Will you be the speaker again, or will you invite others?
- How can you expand the pool of speakers?
- Where will you get your audience from?
- What should you do if my audience wants to use a webinar software that is different than the one of your choice?

Be open and flexible: webinars may not suit everyone or every topic. There may be distinct advantages to meeting face to face with an audience or developing a more "static" online tutorial to answer frequently asked questions. It is not a matter of which one is best; face-to-face training, tutorial-based e-learning, and webinars all have pros and cons. We hope that these 10 simple rules will be tried, tested, and adjusted as you see fit. Let us know how it goes.

## Conclusion

Webinars are a powerful and engaging means of training and dissemination [32] that can reach global audiences and therefore help us to address inequality and imbalance of teaching bioinformatics or other subjects. How can you deliver a stress-free webinar that can be available to everyone everywhere? Our 10 simple rules will decrease the activation energy of delivering webinars: you will feel better equipped to adopt this alternative method of training and take your subject of expertise to places where face-to-face training is not available. We hope that our experience can inspire you into this brave new world of live distance training. You will then be able to add your own rules, especially if you decide to use webinars in contexts other than training only.

## Acknowledgments

We thank Giulietta Spudich who, in addition to setting up the first webinars at EMBL-EBI in 2013 with Gary Battle, had a pivotal role in many of our webinars and shared her knowledge, enthusiasm, and care for users and webinar attendees. We are also in debt with Gabriella Rustici, Sung Hee Park, Dominic Clark, and the EMBL-EBI Training (both past, namely Richard Grandison, and present members) for their valuable support. A big thank you to all our webinars' attendees and for their time in giving us their feedback, which has helped us to improve our webinars. We would also like to thank the viewers of our recording webinars, now available on YouTube. This article was inspired by discussions with Miguel Pignatelli.

## References


1. Via A, De Las Rivas J, Attwood TK, Landsman D, Brazas MD, Leunissen JAM, et al. Ten Simple Rules for Developing a Short Bioinformatics Training Course. PLoS Comput Biol. 2011 Oct 27; 7(10): e1002245. https://doi.org/10.1371/journal.pcbi.1002245 PMID: 22046119
2. Searls DB. Ten Simple Rules for Online Learning. PLoS Comput Biol. 2012 Sep 13; 8(9):e1002631. https://doi.org/10.1371/journal.pcbi.1002631 PMID: 23028268
3. Kemp N, Grieve R. Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. Front Psychol. 2014 Nov 12; 5:1278. https://doi.org/10.3389/fpsyg.2014.01278 PMID: 25429276
4. Train at EMBL-EBI [Internet]. Available from: https://www.ebi.ac.uk/training/handson. [cited 2018 Oct 30].
5. Training at Bioinformatics Barcelona [Internet]. Bioinformatics Barcelona BIB. Available from: http://www.bioinformaticsbarcelona.eu/training/training-at-bib/. [cited 2018 Sep 7].
6. Canadian Bioinformatics Workshops 2018 [Internet]. bioinformatics.ca. Available from: https://bioinformatics.ca/workshops/. [cited 2018 Jun 29].
7. Events | UC Davis Bioinformatics Core [Internet]. UC Davis Bioinformatics Core. Available from: http://bioinformatics.ucdavis.edu/training/events/. [cited 2018 Jun 29].
8. H3ABioNet. H3ABioNet [Internet]. Pan African Bioinformatics Network for H3Africa. Available from: https://www.h3abionet.org/. [cited 2018 Mar 29].
9. CABANA Project. What is CABANA? | CABANA [Internet]. Capacity building for bioinformatics in Latin America. Available from: www.cabana.online. [cited 2018 Mar 29].
10. Khan AM, Tan TW, Schönbach C, Ranganathan S. APBioNet—Transforming Bioinformatics in the Asia-Pacific Region. PLoS Comput Biol. Public Library of Science; 2013; 9: e1003317. https://doi.org/10.1371/journal.pcbi.1003317 PMID: 24204244
11. EMBL-EBI Train online [Internet]. EMBL-EBI Training | Train online. Available from: https://www.ebi.ac.uk/training/online/. [cited 2018 Jul 9].
12. Attwood TK, Leskosek BL, Dimec J, Morgan S, Mulder N, van Gelder CWG, et al. Defining a lingua franca for the ELIXIR/GOBLET e-learning ecosystem. https://doi.org/10.5281/zenodo.166378
13. The GOBLET training portal: a global repository of bioinformatics training materials, courses and trainers. Available from: http://dx.doi.org/10.1093/bioinformatics/btu601. [cited 2018 Aug 10].
14. EMBL-EBI Webinars [Internet]. EMBL-EBI Training | Webinars. Available from: https://www.ebi.ac.uk/training/webinars?ebi-type=webinar. [cited 2018 Aug 10].
15. De Las Rivas J, Bonavides-Martínez C, Campos-Laborie FJ. Bioinformatics in Latin America and SoIBio impact, a tale of spin-off and expansion around genomes and protein structures. Brief Bioinform. 2017; https://doi.org/10.1093/bib/bbx064 PMID: 28981567
16. Mir S, Alhroub Y, Anyango S, Armstrong DR, Berrisford JM, Clark AR, et al. PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. Nucleic Acids Res. 2018 Jan 4; 46(D1): D486–92. https://doi.org/10.1093/nar/gkx1070 PMID: 29126160
17. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018 Jan 4; 46(D1):D754–61. https://doi.org/10.1093/nar/gkx1098 PMID: 29155950
18. Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, et al. Ensembl 2006. Nucleic Acids Res. 2006 Jan 1; 34(Database issue):D556–61. https://doi.org/10.1093/nar/gkj133 PMID: 16381931
19. Velankar S, Best C, Beuth B, Boutselakis CH, Cobley N, Da Silva Sousa AW, et al. PDBe: Protein Data Bank in Europe. Nucleic Acids Res. 2010 Jan; 38(Database issue):D308–17. https://doi.org/10.1093/nar/gkp916 PMID: 19858099
20. Flicek P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, et al. Ensembl's 10th year. Nucleic Acids Res. 2010 Jan 1; 38(suppl_1):D557–62. https://academic.oup.com/nar/article/38/suppl_1/D557/3112317 PMID: 19906699
21. Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, et al. PDBe: Protein Data Bank in Europe. Nucleic Acids Res. 2014 Jan 1; 42(Database issue):D285–91. https://doi.org/10.1093/nar/gkt1180 PMID: 24288376
22. EMBL-EBI. Industry Program [Internet]. Available from: https://www.ebi.ac.uk/industry. [cited 2018 Feb 28].
23. Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. Ten Simple Rules for Effective Online Outreach. PLoS Comput Biol. 2015 Apr 16; 11(4):e1003906. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439
24. Free Educational Virtual Events: Science & Medicine | LabRoots | On Demand [Internet]. Available from: https://www.labroots.com/virtual-events/all/filter/ondemand/page. [cited 2018 Feb 28].

25. Training sessions—Open Targets Platform Outreach [Internet]. Available from: https://www.targetvalidation.org/outreach. [cited 2018 Oct 30].
26. Webinar on Transcript Support Levels, APPRIS, and Vervet Monkey [Internet]. 2014. Available from: https://www.youtube.com/watch?v=rTlIFCUh47I&list=PLqB8Yx1tGBMbBBBB8yFG9zSblpYBK6zBv&index=8. [cited 2018 Feb 28].
27. Open Targets: Mining gene and disease associations for improved drug target identification [Internet]. 2017. Available from: https://www.youtube.com/watch?v=vH2-8B7JqXE&list=PLncWVtwSXtqb8PyL6-ENSCuqP7_4Aj5BE. [cited 2018 Mar 1].
28. Ensembl Highlights Webinar [Internet]. 2013. Available from: https://www.youtube.com/watch?v=-9e3WPlsk8o&list=PLqB8Yx1tGBMbBBBB8yFG9zSblpYBK6zBv. [cited 2018 Mar 1].
29. How to use EMBL-EBI's freely available databases and tools [Internet]. Available from: http://www.youtube.com/playlist?list=PL67E0627174F36FCF. [cited 2018 Mar 1].
30. Matthews J. Surveys that work: an introduction to using survey methods [Internet]. 2017. Available from: http://www.effortmark.co.uk/surveys-work-introduction-using-survey-methods/. [cited 2018 Feb 28].
31. Jarrett C. Surveys that work EBI_2017 [Internet]. 2017. Available from: https://www.slideshare.net/cjforms/surveys-that-work-ebi2017. [cited 2018 Mar 1].
32. Elixir Events—Webinars [Internet]. Elixir Events | Webinars. Available from: https://www.elixir-europe.org/events?f%5B0%5D=field_type%3A14. [cited 2018 Oct 30].

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules to create a serious game, illustrated with examples from structural biology

**Marc Baaden[1], Olivier Delalande[2], Nicolas Ferey[3], Samuela Pasquali[4], Jérôme Waldispühl[5], Antoine Taly[1]***

**1** Laboratoire de Biochimie Théorique, Université Paris Diderot, Sorbonne Paris Cité, Paris, France, **2** Institut de Génétique et Développement de Rennes, Univ. Rennes 1, Rennes, France, **3** VENISE group, CNRS-LIMSI, Orsay, France, **4** Laboratoire de Cristallographie et RMN Biologiques, Faculté des sciences pharmaceutiques et biologiques, Université Paris Descartes et Université Sorbonne Paris Cité, Paris, France, **5** School of Computer Science, McGill University, Montréal, Canada

* taly@ibpc.fr





**Citation:** Baaden M, Delalande O, Ferey N, Pasquali S, Waldispühl J, Taly A (2018) Ten simple rules to create a serious game, illustrated with examples from structural biology. PLoS Comput Biol 14(3): e1005955. https://doi.org/10.1371/journal.pcbi.1005955

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** March 8, 2018



**Funding:** This work was supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Overview

Serious scientific games are games whose purpose is not just fun. In the field of science, the serious goals include crucial activities for scientists: outreach, teaching and research. The number of serious games is increasing rapidly, in particular citizen science games (CSGs), games that allow people to produce and/or analyze scientific data. It is possible to build a set of rules providing a guideline to create or improve serious games. We present arguments gathered from our own experience (Phylo, DocMolecules, the HiRE-RNA contest and Pangu) as well as examples from the growing literature on scientific serious games.

## Introduction

Science has an enormous impact on society; therefore, understanding and participating in science projects is important for citizens. The involvement of nonscientists in the realization and design of science projects is called citizen science. Citizen science has become more abundant during the past decade [1]. One good example of a citizen science project is GalaxyZoo, which engages participants in classifying galaxies and has produced numerous publications—48 by 2014 [1].

This way of developing a research program is on the rise; one striking example is the game Foldit [2–3]. The Foldit project is an online 3D jigsaw puzzle in which players are invited to shake and wiggle the 3D structure of proteins to find their most stable conformations [4]. Since its release in 2008, the project has brought remarkable results from a biological point of view [2–3], but it was also useful to collaboratively develop new algorithms to solve a particular scientific problem [5]. Indeed, a very recent study showed that the results of players for model refinement tasks can be compared favorably with those of professional researchers [6]. Similarly, Mazzanti et al. developed the HiRE-RNA contest and showed that novice players are able to fold RNA structures without much prior knowledge [7].

The success of citizen science initiatives is at least partly related to the ability of groups to perform many tasks better than the sum of their individuals, the "wisdom of crowds effect." Many studies have gathered information on the determinants of collective intelligence [8], particularly thanks to controlled experiments in "crowd wisdom" [9]. These studies have shown that key variables need to be scrutinized, such as information network structure [10], communication between users [11], and social influence [12]. Based on these fundamental observations and our own experience, we present 10 simple rules to create or improve citizen science games (CSGs) in this developing field. We focus on the computational molecular biology area, in which CSGs are especially frequent. We share experience from our own Phylo, DocMolecules, HiRE-RNA, and Pangu projects and compare it to several other initiatives such as FoldIt and EteRNA. Such games may fall into multiple categories such as collecting scientific data, sorting it, or solving problems.

## Rule 1: Define a (serious) goal

The most synthetic definition of a serious game is that of video game designers Sande Chen and David Michael: a game "in which education (in its various forms) is the primary goal rather than entertainment" [13]. The work of Julian Alvarez, Damien Djaouti, and Olivier Rampnoux [14] further defined a serious game as a device, digital or otherwise, whose initial intention is to consistently combine utilitarian aspects with playful means. Such an association is aimed at an activity or a market other than entertainment alone. Therefore, a purpose needs to be clearly defined in terms of science, outreach, and teaching. A good game may address all three aspects. Knowing that professional video game production can cost millions of dollars, funding should also be taken into consideration from the very beginning of the project. The funding impacts all aspects of the project. Thus, having a clear idea helps designers to be realistic about the goals that can be achieved.

### Is it to produce scientific data?

Many CSGs have a simulation component that allows the players to interact with and/or produce scientific data. Therefore, CSGs should lead to discoveries that can ultimately be published in the scientific literature. The scientific relevance of the results of gameplay increases the player’s interest and motivation.

We note that up to now, most of the publications on the various CSGs mainly concern the games themselves, discussing the quality of data generated, impact on motivation, etc. This outcome is expected, as most of the initial publications were proof of principle. So far, only a few games generated actual scientific results on the subject they were meant to study. One intriguing common point among the first projects that published data or results obtained using data generated by players on open-ended questions (e.g., Phylo, Foldit, Eyewire [15], and EteRNA) is that they all involved pattern-matching tasks. However, although serious games date back to before the 1980s, such games with a scientific twist are relatively young, such that conclusions are difficult to make.

### Is it outreach?

There is only a small step from a citizen science project to outreach, because the involvement of participants is a criterion for success in citizen science projects [16]. This natural link leads to adaptation opportunities. Outreach can be an objective per se, as in the case of DocMolecules, which uses simulation and visualization tools developed for other projects on interactive docking to convey the molecular-level action of a drug in the fight against allergies.

### Is it teaching?

Videogames have properties that make them adequate learning platforms [17]. Games developed originally for research are regularly used for teaching as well. Good examples are Foldit [18], Phylo [19], EteRNA [20], and the HiRE-RNA contest [7]. The difference between research and education use lies partially in the terms used, i.e., in-game actions are sometimes called by a name that is presumably more familiar to the players but that hides the correct scientific term. Unfortunately, this choice may have consequences for learning that can only be limited with a debriefing to make the right connection between game and course. Therefore, it could be advised not to sacrifice precision of scientific terms when players may use the game for educational purposes (as well as for people primarily interested in science). Another aspect to bear in mind is that the success of informal learning around games depends on players' profiles [21].

## Rule 2: Fine-tune the balance between entertainment and serious tasks

As mentioned above, a serious game is a chimera of a utilitarian goal and game mechanics. Ideally, the game design should be implemented as a function of the objectives of the game (e.g., data production, knowledge diffusion). Equilibrium and compromise need to be found between scientific accuracy and player accessibility ([22], p51). The tradeoff particularly applies to 1) visualization and graphics, 2) interaction design, and 3) scoring [4].

The level of simplification of scientific information is a key point. Not all players are looking for the same level of information. Therefore, providing access to more advanced material can help to keep experts around. For example, Phylo integrates an expert interface accessible to users who play at least 20 games with the classic edition, which allows them to play on larger grids (300 columns) than those used in the basic version (25 columns). This feature helps to increase the engagement of the most assiduous users.

Entertainment can also be used as a reward for achievements in the game. For example, short animated sequences related to the scientific topic can be both informative and entertaining at the end of a completed level in the game.

## Rule 3: Enable the player to interact with scientific data

Use of scientific data enables game designers to raise player interest and aim for participative data production of high scientific relevance. Intuitive interaction with the data (e.g., through molecular simulations) enriches the learning experience. One way to allow interaction with scientific data is to derive a "single quantitative metric of success" that facilitates the transformation of a task into a game element [23]. For instance, Phylo uses a simplified scoring function that estimates the quality of an alignment.

One route to this use of data is to recycle available simulation tools. The prototypical example is Foldit, which derives from the Rosetta@home program [24], which in turn builds on the Rosetta software [25]. Similarly, the software UnityMol [26] and BioSpring [27] were used to create DocMolecules. DocMolecules uses Protein Data Bank (PDB) structures as input for biological targets and molecular models for drugs. Use of force field terms such as nonbonded interactions and precomputed electrostatic fields drives feedback loops in the game.

When generating data, the considerations presented in "Ten Simple Rules for Effective Online Outreach" [28] apply, especially Rule #8 ("Collect and assess data"). An interesting aspect of CSGs, in which the data are generated by volunteers, is whether the data are made available. There is a general tendency to increase the availability of the data. For example, metrics measured in the game, and largely used for the development of the game itself, can be made open [29].

The availability of the data generated by the players paves the way for another change in perspective for the participants. In addition to receiving the data they generated predigested by professional scientists, players can also access untreated information that allows them to analyze these data. A striking example is the resolution of a previously unsolved crystal structure by scientists thanks to a Foldit-generated model [30]. Furthermore, the source code of some CSGs has been publicly released to foster the collective development of these platforms (e.g., Phylo, Mark2Cure).

## Rule 4: Promote onboarding and engagement

Players have heterogeneous expectations [31]. Therefore, the reward system should be versatile. The entry barrier should be low, and ideally, the difficulty is adapted to each player ([22], p49, 52). For example, the background of players (general public versus students) has been found to be related to players' abilities [32]. Defining player activities (gameplay) should build on tasks that the participants enjoy doing/completing ([22], p53).

It is often proposed that adding game components to a serious task should increase the motivation of users. However, it can also be argued that using this extrinsic motivation could interfere with the user's intrinsic motivation. In the case of CSGs, this could lead to alteration of the data generated by players. A recent study suggests that this concern is not necessarily justified [33].

It is crucial to provide feedback to the players about their progress. This can be done through competitive score production (e.g., best scores tables, online community), which may create a knock-on effect on players who will progress more quickly in knowledge and competence acquisition. Game mechanics can orient player contributions to allow covering a project need (as when Foldit encouraged solving many targets in the Critical Assessment of protein Structure Prediction [CASP] competition, a worldwide competition allowing laboratories to test their folding methods [34]). The overall aim is to increase player engagement. However, it should be kept in mind that player profiles favoring competition are only a fraction of the population. Nevertheless, for CSGs, as for other crowdsourcing programs, the results rely mainly on a few participants that contribute most of the data, called "whales" [35].

It should be kept in mind that the motivations of players are not those of professional scientists. For example, Foldit players offered the possibility to be coauthors of articles decided rather to be credited together with their Foldit team [23].

## Rule 5: Manage information flow

The information given to individuals in a group has an important impact on their collective behavior, i.e., the information to which they have access might push them to copy others, which could lead to specific behavior [11]. Exchange of information between participants, as well as between the system (the serious game and its backend) and participants, is a crucial point [23]. However, this dialogue can hardly be established solely by scores, which provide rather limited information and can even be misleading [7]. The exchange of data is part of what allows collective intelligence to emerge. Indeed, the ability to communicate has a significant impact on the emergence of collective intelligence [8], with the network structure having a strong influence [9]. This is illustrated by the study of Tinati et al., which noted positive effects of communication in web-based citizen science programs ([36], see also references cited therein). A good example is found in EteRNA, in which players up-vote their favorite solution.

Furthermore, specific behaviors were observed, allowing participants to be classified in categories: discoverer, hypothesizer, and investigator [37]. When a game allows discussions between members of the community, another category of participants mainly interested in using this media for socializing can also participate in the diffusion of ideas within the crowd. These behaviors can lead to collective intelligence. So far, CSGs with active forums tend to be those that produce articles on application results generated by participants.

Allowing communication between users can create bias and be at the origin of group thinking phenomena. This could be due to complementary effects such as social influence effects, the range reduction effect, and the confidence effect [12]. To avoid this, Amazon Mechanical Turk deliberately forbids communication between participants. On the other hand, allowing users to exchange information opens the door to more complex calculations [38].

## Rule 6: Provide an appropriate narrative

The narrative is an important aspect in many games [39]. This is also true for serious games: "Stories are equally important for serious and non-serious games alike" [40]. The plot should give sense and context to the game so that everything is connected coherently and the player knows his/her part in the overall story: "While we cannot always control the actions of the player or the way she plays the game, we can adjust our storytelling technique to better align our learning objectives with our dramatic objectives" [40]. A storyboard can prove helpful for development, particularly for games that must reconcile divergent serious and game objectives [41].

To better target the audience (depending on the project), it is necessary to convey the difficulty of progressing or switching between different levels. An introductory tutorial sequence helps to provide a progression in the dispensing of scientific information (both concepts and vocabulary). With DocMolecules, this is done through two separate levels of the game: the first level presents the context and cellular aspects, and the second level allows players to manipulate the drug molecule in order to dock it. However, the Foldit team has noticed that a tutorial may not be mandatory when the user can discover the rules through experimentation [42].

## Rule 7: Adapt your level design

Depending on the objective and the audience, the degree of simplification of scientific content and manipulation must be adjusted, which is essential to make the game accessible to a broad audience and maintain player interest [22].

Game duration needs to be adapted. The serious goal implies that the game needs to be adapted to everyone, including those who define themselves as non-gamers, which in turn implies to develop casual games. Casual games are easily accessed, simple to control, and not punishing [22].

In order to control the difficulty, ideally the gameplay should not exceed 5–20 minutes, and the game should be targeted to the right person, who will find it rewarding [22]. It is therefore important to be able to predict a task's difficulty, which is essential for channeling these tasks to players with the appropriate skill level [32].

## Rule 8: Develop good graphics, not just for eye candy

The primary scientific data is often complex; adapted graphical representations will help significantly to better understand it (Fig 1). High-quality graphics increase the player's immersion. In the context of displaying complex scientific objects such as molecules for example, shadowing is a necessary feature for volume rendering and shape perception. Adjusting the crowding of the game scene allows realistic rendering of molecular worlds [43]. Indeed, experience from scientific molecular movies indicates that there are several ways to represent biological molecules and their crowding [44].

In representing molecules, the game could switch between ball-and-stick and molecular surfaces depending on the size and nature of the molecules to be displayed (e.g., to distinguish small chemical compounds from large biological macromolecules). Hiding information

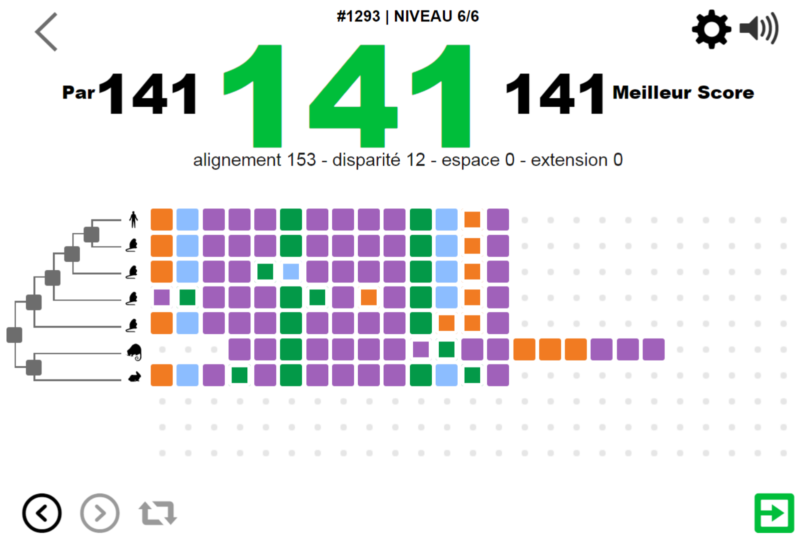



**Fig 1. Screenshot of Phylo illustrating the high level of information that can be provided to the player through graphics.**

https://doi.org/10.1371/journal.pcbi.1005955.g001

depending on player activities can also be used, as in the case of toggling on and off the side chain representation in Foldit [2].

## Rule 9: Use all modalities, particularly sound

The data a player will have to deal with in a serious game may be very complex and multidimensional [45]. If everything is conveyed visually, that channel may quickly become overloaded, and the player will be lost. One solution is to make use of several channels, not only the visual one, to convey important properties. Sound effects and music can drive interest or increase scenario effects. A good option is to simplify the rendering by conveying some information through sound rather than visual effects (e.g., score progress or formation of an interaction) [46].

Another modality that can be developed is touch (for example, through manipulation by hand with augmented reality applications) [47]. An example is presented in our Pangu project (Fig 2).

## Rule 10: "Iteratively assess what works and what doesn't"

To illustrate how important it is to prototype, evaluate, and iterate, we use here a title taken from the article "Ten Simple Rules for Effective Online Outreach" [28], in which this is Rule #9.

It is common practice to do iterations in game development [48], but in the case of serious games, this process involves three groups of evaluators instead of two. In addition to players and developers, scientists must come into play [4]. An interesting conclusion made by Cooper et al. [4] is that "we have also learned not to expect the way that expert scientists view the problem to be the best way for players." The iteration process can address many of the points described above, such as visualization and graphics, interaction design, and scoring mechanisms.

After creating a prototype, a critical step is to test it, which requires defining evaluation criteria for both game components and utilitarian aspects. The criteria defined to study the

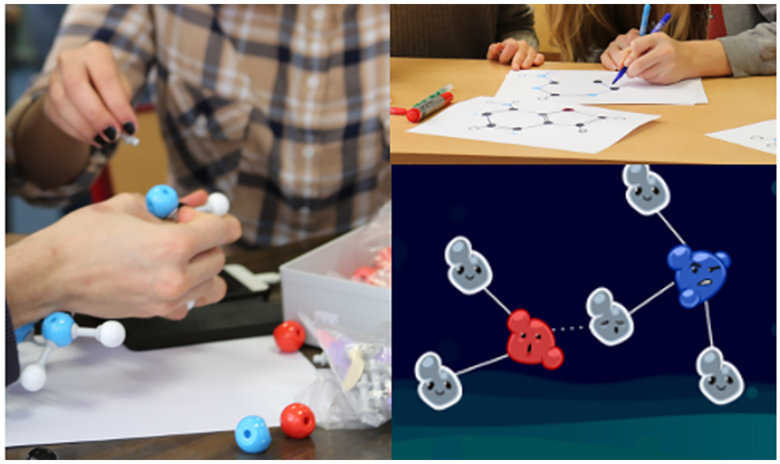


**Fig 2. Example of an augmented reality application (Pangu).** Players build their tangible model of the expected molecule with a standard molecular model kit (or draw it on paper) and scan it with a mobile device, which validates that the correct molecule was built.

https://doi.org/10.1371/journal.pcbi.1005955.g002

citizen science project Zooniverse, i.e., a "success matrix" measuring both contribution to science and public engagement, could be used as an example [16].

The final step is to iterate when appropriate. It seems natural that a CSG uses players to assess quality. Players may even become valuable contributors and propose original game content, as seen with Open-Phylo [49].

## Conclusion

The area of serious games has a long history, almost as long as video games themselves [50]. Recently, technological opportunities, including the internet, have allowed for the rapid expansion of CSGs [51]. This development represents a great opportunity by itself and should find more applications in the future with the democratization of virtual and augmented reality. Yet designing a good game remains a tricky business with many pitfalls. We hope that the guidelines provided above will help any scientific game designer to achieve successful implementation of a scientific endeavor within game mechanics.

## References


1. Follett R, & Strezov V (2015). An analysis of citizen science based research: usage and publication patterns. PLoS ONE, 10(11), e0143687. https://doi.org/10.1371/journal.pone.0143687 PMID: 26600041
2. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. (2010). Predicting protein structures with a multiplayer online game. Nature, 466(7307), 756–760. https://doi.org/10.1038/nature09304 PMID: 20686574
3. Good B M, & Su A I (2011). Games with a scientific purpose. Genome biology, 12(12), 135. https://doi.org/10.1186/gb-2011-12-12-135 PMID: 22204700
4. Cooper, S, Treuille, A, Barbero, J, Leaver-Fay, A, Tuite, K, Khatib, F., et al. The challenge of designing scientific discovery games. In Proceedings of the Fifth international Conference on the Foundations of Digital Games (pp. 40–47). ACM.
5. Khatib F, Cooper S, Tyka M D, Xu K, Makedon I, Popović Z, et al. (2011). Algorithm discovery by protein folding game players. Proceedings of the National Academy of Sciences, 108(47), 18949–18953.
6. Horowitz S, Koepnick B, Martin R, Tymieniecki A, Winburn AA, Cooper S, et al., University of Michigan students, Popović Z., Baker D., Khatib F., & Bardwell J.C.A. (2016). Determining crystal structures through crowdsourcing and coursework. Nature communications, 7, 12549. https://doi.org/10.1038/ncomms12549 PMID: 27633552

7. Mazzanti L, Doutreligne S, Gageat C, Derreumaux P, Taly A & Baaden M. What can human-guided simulations bring to RNA folding? Biophysical Journal 113 (2) 302–312. https://doi.org/10.1016/j.bpj.2017.05.047 PMID: 28648754
8. Woolley AW, Chabris CF, Pentland A, Hashmi N, & Malone TW (2010). Evidence for a collective intelligence factor in the performance of human groups. Science, 330(6004), 686–688. https://doi.org/10.1126/science.1193147 PMID: 20929725
9. Centola D, & Baronchelli A (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. Proceedings of the National Academy of Sciences, 112(7), 1989–1994.
10. Evans JA, & Foster JG (2011). Metaknowledge. Science, 331(6018), 721–725. https://doi.org/10.1126/science.1201765 PMID: 21311014
11. Laland KN (2004). Social learning strategies. Learning & behavior, 32(1), 4–14.
12. Lorenz J, Rauhut H, Schweitzer F, & Helbing D (2011). How social influence can undermine the wisdom of crowd effect. Proceedings of the National Academy of Sciences, 108(22), 9020–9025.
13. Michael D, Chen S (2005), Serious games that educate train and inform, Course Technology PTR, USA
14. Alvarez J, Djaouti D, & Rampnoux O (2016). Apprendre avec les serious games? Poitiers: Canopé Edition.
15. Kim JS, Greene MJ, Zlateski A, Lee K, Richardson M, Turaga SC, et al. (2014). Space-time wiring specificity supports direction selectivity in the retina. Nature, 509(7500), 331–336. https://doi.org/10.1038/nature13240 PMID: 24805243
16. Graham GG, Cox J, Simmons B, Lintott C, Masters K, Greenhill A, et al. (2015) How is success defined and measured in online citizen science: a case study of Zooniverse projects. Computing in science and engineering, PP (99) (22). ISSN 1521-9615 https://doi.org/10.1109/MCSE.2010.27
17. Gee JP (2003). What video games have to teach us about learning and literacy. Computers in Entertainment (CIE), 1(1), 20–20.
18. Franco J (2012). Online gaming for understanding folding, interactions, and structure. Journal of Chemical Education, 89(12), 1543–1546.
19. Kawrykow A, Roumanis G, Kam A, Kwak D, Leung C, Wu C, et al. (2012). Phylo: a citizen science approach for improving multiple sequence alignment. PLoS ONE, 7(3), e31362. https://doi.org/10.1371/journal.pone.0031362 PMID: 22412834
20. Lee J, Kladwang W, Lee M, Cantu D, Azizyan M, Kim H, et al. (2014). RNA design rules from a massive open laboratory. Proceedings of the National Academy of Sciences, 111(6), 2122–2127.
21. Iacovides I, McAndrew P, Scanlon E, & Aczel J (2014). The gaming involvement and informal learning framework. Simulation & Gaming, 45(4–5), 611–626.
22. Law E, & Ahn LV (2011). Human computation. Synthesis Lectures on Artificial Intelligence and Machine Learning, 5(3), 1–121.
23. Cooper S, Khatib F, & Baker D (2013). Increasing public involvement in structural biology. Structure, 21 (9), 1482–1484. https://doi.org/10.1016/j.str.2013.08.009 PMID: 24010706
24. Das R, Qian B, Raman S, Vernon R, Thompson J, Bradley P.,et al.(2007). Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@ home. Proteins: Structure, Function, and Bioinformatics, 69(S8), 118–128.
25. Das R, & Baker D (2008). Macromolecular modeling with rosetta. Annu. Rev. Biochem., 77, 363–382. https://doi.org/10.1146/annurev.biochem.77.062906.171838 PMID: 18410248
26. Doutreligne S, Gageat C, Cragnolini T, Taly A, Pasquali S, Baaden M, et al. UnityMol: Simulation et Visualisation Interactive à des fins d'Enseignement et de Recherche. GGMM 2015, 118.
27. Ferey N, Delalande O, & Baaden M (2012). Biospring: an interactive and multi-resolution software for flexible docking and for mechanical exploration of large biomolecular assemblies. JOBIM 2012-Journées Ouvertes en Biologie, Informatique et Mathématiques, 433–434.
28. Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. (2015) Ten Simple Rules for Effective Online Outreach. PLoS Comput Biol 11(4): e1003906. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439
29. Himmelstein J, Goujet R, Duong TK, Bland J, Lindner AB. Human Computation (2016) 3:1:119–141 2016.
30. Gilski M, Kazmierczyk M, Krzywda S, Zábranská H, Cooper S, Popović Z, et al. (2011). High-resolution structure of a retroviral protease folded as a monomer. Acta Crystallographica Section D: Biological Crystallography, 67(11), 907–914.
31. Bartle, RA (1990). Who Plays MUAs? Comms Plus!, October/November 1990 18–19.

32. ] Singh, Ahsan, Blanchette, Waldispühl (To Appear). Lessons from an online massive genomics computer game. In Proceedings of the 5th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2017).

33. Prestopnik N, Crowston Kevin, Wang Jun, Gamers, citizen scientists, and data: Exploring participant contributions in two games with a purpose, Computers in Human Behavior, Volume 68, March 2017, Pages 254–268.

34. Moult J, Fidelis K, Kryshtafovych A, Schwede T, & Tramontano A (2017). Critical assessment of methods of protein structure prediction (CASP)—Round XII. *Proteins*: *Structure*, *Function*, *and Bioinformatics*.

35. Sauermann H, & Franzoni C (2015). Crowd science user contribution patterns and their implications. Proceedings of the National Academy of Sciences, 112(3), 679–684.

36. Tinati, R, Simperl, E, & Luczak-Roesch, M (2017). To help or hinder: Real-time chat in citizen science.

37. Tinati, R, Simperl, E, Luczak-Roesch, M, Van Kleek, M, & Shadbolt, N (2014). Collective Intelligence in Citizen Science—A Study of Performers and Talkers. arXiv preprint arXiv:1406.7551.

38. Tremblay-Savard, O, Butyaev, A and J Waldispühl. 2016. Collaborative Solving in a Human Computing Game Using a Market, Skills and Challenges. In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16). ACM, New York, NY, USA, 130–141

39. Jenkins H (2004). Game design as narrative architecture. Computer, 44, 53

40. McDaniel R, Fiore SM, & Nicholson D (2010). Serious storytelling: Narrative considerations for serious games researchers and developers. In Serious game design and development: Technologies for training and learning (pp. 13–30). IGI Global.

41. Belanich J, Orvis KB, Horn DB, & Solberg JL (2011). Bridging game development and instructional design. In Instructional Design: Concepts, Methodologies, Tools and Applications (pp. 464–479). IGI Global.

42. Andersen, E, O'Rourke, E, Liu, YE, Snider, R, Lowdermilk, J, Truong, D, et al. (2012, May). The impact of tutorials on games of varying complexity. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 59–68). ACM.

43. Ellis RJ (2001). Macromolecular crowding: obvious but underappreciated. Trends in biochemical sciences, 26 (10), 597–604. PMID: 11590012

44. McGill G (2008). Molecular movies. . . coming to a lecture near you. Cell, 133(7), 1127–1132. https://doi.org/10.1016/j.cell.2008.06.013 PMID: 18585343

45. O'Donoghue SI, Gavin AC, Gehlenborg N, Goodsell DS, Hériché JK, Nielsen CB, et al. (2010). Visualizing biological data—now and in the future. Nature methods, 7, S2–S4. https://doi.org/10.1038/nmeth.f.301 PMID: 20195254

46. Ballweg, H, Bronowska, AK, & Vickers, P (2017). Interactive Sonification for Structural Biology and Structre-based Drug Design.

47. Férey N, Nelson J, Martin C, Picinali L, Bouyer G, Tek A, et al. (2009). Multisensory VR interaction for protein-docking in the CoRSAIRe project. Virtual Reality, 13(4), 273.

48. O'Hagan AO, & O'Connor RV (2015). Towards an Understanding of Game Software Development Processes: A Case Study. In European Conference on Software Process Improvement (pp. 3–16). Springer International Publishing.

49. Kwak D, Kam A, Becerra D, Zhou Q, Hops A, Zarour E, et al. (2013). Open-Phylo: a customizable crowd-computing platform for multiple sequence alignment. Genome biology, 14(10), R116. https://doi.org/10.1186/gb-2013-14-10-r116 PMID: 24148814

50. Djaouti D, Alvarez J, Jessel JP, & Rampnoux O (2011). Origins of serious games. In Serious games and edutainment applications (pp. 25–43). Springer London.

51. Curtis V (2014). Online citizen science games: opportunities for the biological sciences. Applied & translational genomics, 3(4), 90–94.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for collaborative lesson development

**Gabriel A. Devenyi[1☯], Rémi Emonet[2☯], Rayna M. Harris[3☯], Kate L. Hertweck[4☯], Damien Irving[5☯], Ian Milligan[6☯], Greg Wilson[7☯]***

**1** Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada, **2** Univ Lyon, UJM-Saint-Etienne, Saint-Étienne, France, **3** The University of Texas at Austin, Austin, Texas, United States of America, **4** The University of Texas at Tyler, Tyler, Texas, United States of America, **5** CSIRO Oceans and Atmosphere, Hobart, Tasmania, Australia, **6** University of Waterloo, Waterloo, Ontario, Canada, **7** DataCamp, Toronto, Ontario, Canada

☯ These authors contributed equally to this work.
* gvwilson@third-bit.com



OPEN ACCESS

**Citation:** Devenyi GA, Emonet R, Harris RM, Hertweck KL, Irving D, Milligan I, et al. (2018) Ten simple rules for collaborative lesson development. PLoS Comput Biol 14(3): e1005963. https://doi.org/10.1371/journal.pcbi.1005963

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** March 1, 2018



**Funding:** The authors received no specific funding for this article.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Lessons take significant effort to build and even more to maintain. Most academics do this work on their own, but leveraging a community approach can make educational resource development more sustainable, robust, and responsive. Treating lessons as a community resource to be updated, adapted, and improved incrementally can free up valuable time while increasing quality.

Despite the success of openness in software development and the curation of Wikipedia, it is an uncommon approach in the academic instructional setting. Each year, thousands of university lecturers teach subjects ranging from first year biology to graduate-level courses in Indian film. Some use textbooks written by one or a few authors, but beyond that they develop and maintain their course materials in isolation.

Given that academic research often depends on sharing, this differing approach to developing pedagogical materials is interesting, but the sociology and psychology behind such a blind spot are beyond the scope of this paper.

The authors have many years of experience with community-developed lessons in the context of research computing in the sciences and humanities through organizations like Software Carpentry and Programming Historian [1]. Software Carpentry was founded in 1998 to teach scientists basic computing skills and has since spawned two sibling organizations called Data Carpentry and Library Carpentry. Programming Historian was founded in 2008 and has evolved into a collaboratively edited site providing lessons to humanities scholars. Their guiding principles are that lessons should be 1) open and easily accessible as well as 2) continually maintained, refined, and improved by a community of contributors.

All open education projects (e.g., massive open online courses) satisfy the first criterion by definition, but very few satisfy the second. In other words, while it is common for open education projects to be occasionally updated by an individual or small team (as happens when a new edition of a book is edited and published), this is not the same as continuous improvement by a large community of contributors. The 10 simple rules that follow summarize what we have learned about doing so as maintainers, editors, and reviewers of lessons used by tens of thousands of people (Fig 1). By following these rules, we contend that it is possible to create higher quality lessons than could be created by an individual or small team, both in terms of accuracy and pedagogy (Fig 2). As an added bonus, the lessons are always up-to-date and require less time per author to develop and maintain.

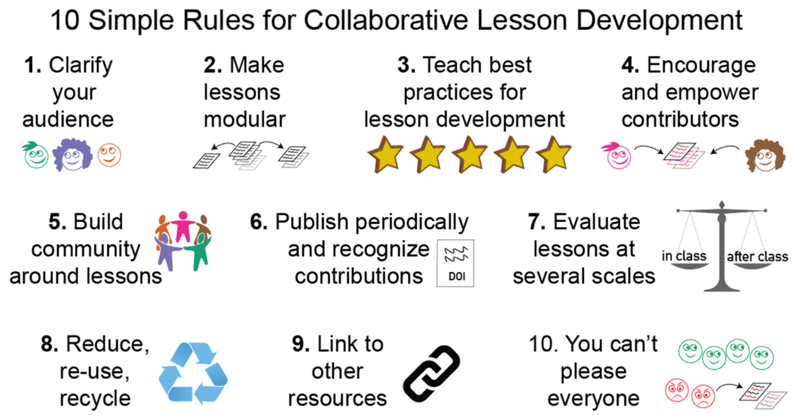



**Fig 1. Graphical abstract of 10 simple rules for collaborative lesson development. 1.** To clarify your audience, consider writing learner profiles (Box 1). **2.** Make lessons modular by breaking them into small, single-purpose modules. **3.** Teach your instructors the best practices for developing, delivering, and maintaining lessons. **4.** Encourage and empower contributors by making the contribution process transparent and straightforward. **5** Build a community around lessons by creating opportunities for participation and mentorship. **6.** Publish new versions periodically and recognize contributors by their unique identifiers (e.g., ORCID). **7.** Evaluate lessons during and after class for a complete picture of their efficacy. **8.** Reduce, reuse, or recycle lessons before creating a new one from scratch. **9.** Link to other resources that complement the lesson content. **10.** Remember that you can't please everyone in your audience or community.

https://doi.org/10.1371/journal.pcbi.1005963.g001

## Rule 1: Clarify your audience

The first requirement for building lessons together is to know for whom they are being built. "Archaeology students" is far too vague: are you and your collaborators thinking of first year

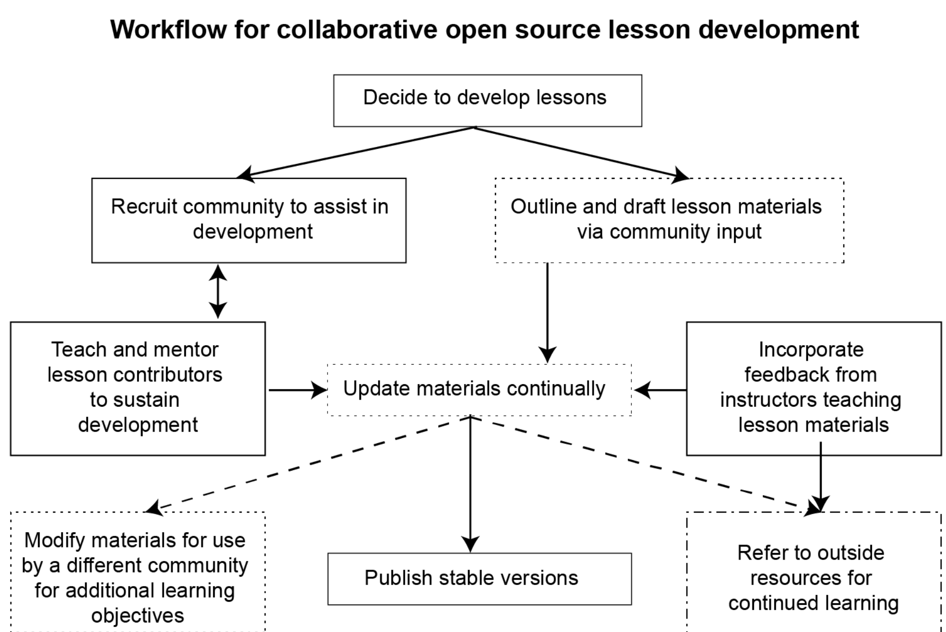



**Fig 2. Collaborative open lesson development.** Following the decision to develop lessons, activities focus on lesson development as well as community building. Boxes surrounded by dotted lines represent community contributions to lessons. Dashed arrows represent connections to activities outside the original lesson design. The box enclosed in a dashed dotted line represents unaffiliated learning resources.

https://doi.org/10.1371/journal.pcbi.1005963.g002

### Box 1: Learner profile.

Jorge has just moved from Costa Rica to Canada to study agricultural engineering. While fluent in both Spanish and English, he has a hearing disability that sometimes makes it hard for him to understand lectures, particularly in noisy environments. Other than using Excel, Word, and the internet, Jorge's most significant previous experience with computers is helping his sister build a WordPress site for the family business.

Jorge needs to measure properties of soil from nearby farms using a handheld device that sends text files to his computer. Right now, Jorge has to open each file in Excel, crop the first and last points, and calculate an average. This workshop will show Jorge how to write a small Python program to read the data, select the right values from each file, and calculate the required statistics.

students who need an introduction to the field, graduate students who intend to specialize in the subdiscipline that is the lesson's focus, or someone in between? If different contributors believe different things about prerequisite knowledge, equipment or software required, or how much time learners will have, they will find it difficult to work together.

Instead of starting with learning objectives (Rule 3), it can be helpful to write learner profiles to clarify the learner's general background, what they already know, what they think they want to do, how the material will help them, and any special needs they might have. This technique is borrowed from user interface design, and a typical learner profile is presented in Box 1.

## Rule 2: Make lessons modular

Every instructor's needs are different, so build small chunks that can be repurposed in many ways. A university lecturer in meteorology, for instance, might construct a course for their students by bringing together lessons on differential equations, fluid mechanics, and absorption spectroscopy. Creating courses this way shifts the instructor's burden from writing to finding and synthesizing, which are easier if lessons clearly define what they cover (Rule 1) and if lessons have been designed by people with a shared worldview (Rule 3).

One way to achieve this is to take existing courses and break them down into smaller, single-purpose modules (a change which has pedagogical and administrative advantages in its own right). When this is done, these modules can be made more discoverable by referencing specific points in the model curricula promulgated by many professional societies (e.g., as learning objectives). Smaller modules are also more approachable for new contributors (Rule 4).

## Rule 3: Teach best practices for lesson development

Decades of pedagogical research have yielded many insights into how best to build and deliver lessons [2]. Unfortunately, many college and university faculty have little or no formal training in education [3], so this knowledge is rarely applied in the classroom.

Our experience is that even a brief introduction to a few key practices helps collaborative lesson development. If people have a shared understanding of how lessons should be developed, it is easier for them to work together. Less obviously, if people have a shared model of how lessons will be used, they are more likely to build reusable material. Finally, teaching people how to teach is a great way to introduce them to each other and build community (Fig 2).

By way of example, Software Carpentry encourages its volunteers to use the popular lesson development methodology presented by Wiggins and McTighe [4], in which learning objectives and assessments are created before any lesson materials are developed. In particular, “summative assessments” are created to determine whether the learning objectives have been met, and “formative assessments” are created to gauge the progress of learners and to give them a chance to practice key skills. The completed formative assessments are put in order and only then are the lessons written, with the aim of connecting each formative assessment to the next. This method is effective in its own right, but its greatest benefit is that it gives everyone a framework for collaboration.

An example of how to teach such pedagogical practices is Software Carpentry’s instructor training program. First offered in 2012, it is now a two-day course delivered both in person and online [5–7]. In addition to a focus on pedagogy, the course teaches whom Software Carpentry’s lessons are for, how they are delivered, and how they are maintained. Largely as a result of this training, several hundred people per year now contribute to Software Carpentry’s lessons.

## Rule 4: Encourage and empower contributors

Making the process for contributing to a lesson simple and transparent is the key to receiving contributions. Licensing, code of conduct, governance, and the review and publication process must all be explicit rather than implicit to lower the social barriers to contribution.

Tools can help, especially if they allow proposed changes to be viewed and discussed prior to their incorporation into the lessons. (In software development, this is known as “premerge review.”) However, some tools that are popular in open-source software development have considerable up-front learning costs. Portals like GitHub, for example, support everything that open lesson development needs but require contributors to use Git, which has a notoriously steep learning curve [8].

Complicating matters further, some file formats make collaboration easier or more difficult. Despite their ubiquity, open-source version control systems do not directly support review or merge of Microsoft Office or OpenDocument file formats, which raises an additional burden for newcomers [9]. While Google Docs and wikis lack some capabilities, such as full-fledged premerge review (although "suggest mode" mitigates this to some degree), their low barrier to entry makes them more welcoming to newcomers.

The best way to choose tools for managing lessons is to ask potential contributors what they are comfortable with rather than requiring them to come to you. Remember also that contributing to a lesson is probably not their top priority, and look for ways to reduce their cognitive load. For example, threaded discussion forums can improve the signal-to-noise ratio by reducing long “reply all” email exchanges. Several open frameworks are available to facilitate development of new lessons, such as learnr (https://rstudio.github.io/learnr), Morea (https://morea-framework.github.io), and DataCamp’s templates (https://www.datacamp.com/teach/documentation).

## Rule 5: Build community around lessons

Software versions and dependencies are constantly changing, while the academic literature is advancing at an ever-increasing pace. As a result, what is cutting edge one year may be out of date the next and simply wrong the year after. Collaborative lesson development groups must therefore focus on creating a community in which contributors support each other rather than relying on a small group of stewards. Authors cannot be expected to maintain continual vigilance on a lesson, but this is necessary for continual use.

A key part of doing this is to create opportunities for legitimate peripheral participation. Curating a list of small tasks that newcomers can easily tackle, encouraging them to give feedback on proposed changes, or asking them to add new exercises and tweak diagrams and references can all provide an on-ramp for people who might question their own authority or ability to change the main body of a lesson. Equally, acknowledging all contributions, however small, gives new contributors an early reward for taking part.

In 2015, Software Carpentry established a Mentoring Subcommittee to support instructors as they progress through training, teaching, and curriculum development. The Mentoring Subcommittee has promoted community building by providing virtual spaces where instructors from all over the world can share success stories and discuss strategies for overcoming challenges. This has helped strengthen the community and provided insight into how lessons can be improved (Rule 7).

Finally, working in the open can be great, but it can also unintentionally suppress voices. Programming Historian makes an ombudsperson available for private chats and facilitation to ensure that no one is excluded. Software Carpentry operates by a Code of Conduct that outlines acceptable standards of behavior for community members and those interacting with the Carpentries at events and in virtual spaces. Community members on a Policy Subcommittee serve as advocates for the Code of Conduct and adjudicate reported violations.

## Rule 6: Publish periodically and recognize contributions

Like software, specific versions of lessons should be published or released periodically so that learners or instructors have something stable to refer to for the duration of their use (Fig 2). Periodic releases also provide an opportunity for recognizing the contributions of new authors and maintainers.

Academia has only a few ways of recognizing contributions. Until these are expanded, it is important to publish lessons in ways that traditional academic systems can digest. One is to give releases DOIs supplied by providers such as Zenodo (https://zenodo.org/) or DataCite (https://www.datacite.org/). Contributors can be listed as authors and the maintainers of the lesson as editors to differentiate recognition of their contributions. Each time the lesson is published, names and identifiers such as ORCIDs (https://orcid.org) should be gathered for all contributors.

A lesson release is a good opportunity to bring the material into a stable shape by fixing outstanding issues and merging contributions. Version control automatically maintains a list of contributors and can also be used to track which content is in which release (e.g., using branches or tags). Lesson releases should use a consistent naming scheme; Software Carpentry has used the year and month of release (e.g., "2017.05") in its releases [10, 11].

If lessons are being released regularly, automate the process and archive old versions in a discoverable location. Also make sure that everyone involved knows what "done" looks like, i.e., which outstanding issues have to be addressed and how they have to be formatted in order for the next release to go out. A simple checklist stored with the lesson materials is good enough to start, but as time goes by, the community may want to use an issue tracking system of some sort so that work items can be assigned to specific people and then ticked off as they are completed.

## Rule 7: Evaluate lessons at several scales

What people immersed in developing lessons think needs fixing can easily differ from what learners think. It is therefore critical to gather and act on feedback at several scales to check assumptions and stay on course (Fig 2).

Microscale feedback can be gathered by an instructor while teaching a particular lesson. Learners can provide feedback on everything from typographical errors and the clarity of quiz questions to the order in which topics are presented, all of which the instructor should record at the end of each class in some shared location (such as a Google Doc or GitHub issues). As well as encouraging direct verbal feedback, it is a good idea to provide learners with a means to provide feedback anonymously during class (e.g., on small pieces of paper like sticky notes or through anonymous surveys).

Surveys and interviews before and after class should be used to uncover larger issues, particularly those arising from developers not fully understanding their audience, e.g., assuming prior knowledge that learners do not have. Such surveys are most effective when conducted 30–90 days after class; this gives people time to reflect, so their feedback will more accurately reflect what they learned rather than how entertained they were. Clearly stated learning objectives (Rule 3) are essential here, as they tell assessors what they should be measuring.

## Rule 8: Reduce, reuse, recycle

Just as a scholar would not write a paper without a literature review, an instructor should not create a new lesson if there is an existing one they could use or contribute to. A short online search can reveal if someone has written what you need, whether it is complementary to your goals, and if it can be tweaked or modified to meet your needs.

Before reusing content, make sure to check its license. Both Programming Historian and the Carpentry projects use the Creative Commons–Attribution license (https://creativecommons.org/licenses/by/4.0/), which allows people to share and adapt material for any purpose as long as they cite the original source. Other Creative Commons licenses may restrict commercial use and/or creation of derivative materials.

The question of licensing also arises when recycling lesson components such as images, data, figures, or code. If the license does not cover them explicitly, ask permission as you would for any other academic material.

The converse of this rule is to make the license for your lessons explicit and discoverable. For example, when lessons are published (Rule 6), make sure that keywords such as "CC-BY" appear in their bibliography entries and HTML page headers.

## Rule 9: Link to other resources

Learners are unlikely to absorb everything they need to know about a topic from your lesson alone. This is partly a matter of scope—any interesting subject is too large to fit in a single lesson—but also a matter of level and direction. As Caulfield has argued [12], the best way to use the internet is to provide a chorus of explanations that offer many angles and approaches for a given topic, each of which may be the best fit for a different set of needs (Fig 2).

Collaboratively developed lessons should direct learners to these resources at strategic points. If a community or discussion forum exists for the topic, such as textbooks, technical documentation, videos, web pages, threads on Quora, or mailing lists, then it is worth including.

Doing this is substantial work, and maintaining it even more so, which makes building community around lessons (Rule 5) all the more important. In particular, it is vital to engage the learners as equal participants in that community. They should be able to propose updates, corrections, and additions to lessons and know that they are encouraged to do so (Rule 4).

## Rule 10: You can't please everyone

No single lesson can be right for every learner. Two people with no prior knowledge of a specific subject may still be able to move at different speeds because of different levels of general background knowledge. Similarly, lessons on ecology for learners in Utah and Vietnam will probably be most relatable if they use different examples. A community may therefore maintain several differently oriented or differently paced lessons on a single topic, just as programming languages provide several different libraries for doing the same general thing with different levels of performance and complexity.

Similarly, no lesson development community can serve all purposes. Some groups may prioritize rapid evolution, while others may prefer a "measure twice, cut once" approach. If there are complementary ways to explain something or points of view that can cohabit respectfully, it may be possible to present them side by side. There are good pedagogical reasons to do this even if contributors do not disagree: weighing alternatives fosters higher-order thinking.

But sometimes choices must be made. The open-source software community has wrestled with these issues for three decades and has evolved some best practices to address them [13]. As discussed in Rule 4, the first step is to have a clear governance structure and a clear, permissive license. Minor disagreements should be discussed openly and respectfully. If they turn out not to be so minor after all, contributors should split off and evolve the lesson in the way they see best. (This is one of the reasons to have a permissive license.)

These splits rarely happen in practice. When they do, it is important to remember that we all share the same vision of better lessons built together.

## Conclusion

Every day, teachers all over the world spend countless hours duplicating each other's work. These 10 rules provide an alternative: adopting the model of collaborative software development to make more robust and sustainable lessons in all domains that can be continually improved by those who use them. We hope that our experiences can help others teach more with more impact and less effort.

## Acknowledgments

We are grateful to everyone who provided feedback on this paper, including James Baker, Nathan Moore, Pariksheet Nanda, Tom Pollard, Byron Smith, and Andrew Walker. We are also grateful to the hundreds of people who have contributed to Programming Historian, Data Carpentry, Software Carpentry, and Library Carpentry over many years.

## References

1. Afanador-Llach MJ, Baker J, Crymble A, Gayol V, Gibbs F, Lincoln M, McDaniel C, Milligan I, Parr J, Castro AR, Knuppel AS, Visconti A, Walsh B, Wieringa J. Programming Historian. Accessed: 8 February 2018. https://programminghistorian.org/.
2. Ambrose SA, Bridges MW, DiPietro M, Lovett MC, Norman MK. How Learning Works. Jossey-Bass; 2010.
3. Brownell S. E. and Tanner K. D.: Barriers to faculty pedagogical change: Lack of training, time, incentives, and. . .tensions with professional identity? CBE Life Sciences Education, 11(4), 2012, https://doi.org/10.1187/cbe.12-09-0163 PMID: 23222828
4. Wiggins G, McTighe J. Understanding by Design. 2nd ed. Association for Supervision and Curriculum Development; 2005.
5. Wilson G. Software Carpentry: Lessons Learned. F1000Research, 3(62), 2016, https://doi.org/10.12688/f1000research.3-62.v2 PMID: 24715981

6. Koch C. and Wilson G. (eds.) Software Carpentry: Instructor Training; 2016. https://zenodo.org/record/57571#.WS8huDOZPdQ.
7. Wilson G. How to Teach Programming (And Other Things). Lulu.com; 2017.
8. GitLab Global Developer Survey. Accessed: 8 February 2018. http://get.gitlab.com/global-developer-survey/, 2016.
9. Jacobs C. T., Gorman G. J., Rees H. E., and Craig L. E.: Experiences With Efficient Methodologies for Teaching Computer Programming to Geoscientists Journal of Geoscience Education 64(3), 2016, https://doi.org/10.5408/15-101.1
10. Devenyi G. A. and Koch C. (eds.): Software Carpentry: The Unix Shell; 2015. https://zenodo.org/record/27355#.WS8lajOZPdQ.
11. Devenyi G. A. and Srinath A. (eds.): Software Carpentry: The Unix Shell; 2017. https://zenodo.org/record/278226#.WS74tTOZPdQ.
12. Caulfield M. Choral Explanations; 2016. Accessed: 8 February 2018. https://hapgood.us/2016/05/13/choral-explanations/.
13. Fogel K. Producing Open Source Software. O'Reilly; 2005.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for developing a mentor–mentee expectations document

**Kristyn S. Masters[1,2,3]*, Pamela K. Kreeger[1,2,4]***

**1** Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI, United States of America, **2** Carbone Cancer Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of America, **3** Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of America, **4** Department of Cell and Regenerative Biology, University of Wisconsin School of Medicine and Public Health, Madison, WI, United States of America

* kmasters@wisc.edu (KSM); kreeger@wisc.edu (PKK)



OPEN ACCESS

**Citation:** Masters KS, Kreeger PK (2017) Ten simple rules for developing a mentor–mentee expectations document. PLoS Comput Biol 13(9): e1005709. https://doi.org/10.1371/journal.pcbi.1005709

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** September 21, 2017



**Funding:** This work was supported by grants from the NIH (1DP2CA195766, PKK; R01GM099031, KSM; R21CA202040, PKK/KSM; R21EY026222, KSM; R01HL093281, KSM; R21EB019508, KSM), NSF (CBET-1401584, KSM/PKK), and the American Cancer Society (RSG-13-026-01-CSM, PKK). The funders had no role in study design, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

There is general agreement that effective mentoring is beneficial for mentees, mentors, and overall scientific productivity [1, 2]. Discussions of what to consider in mentoring philosophies and mentor–mentee relationships have been published [3–5], and discipline-specific versions of a curriculum to develop mentoring skills are available (https://mentoringresources.ictr.wisc.edu). However, these resources focus on general concepts about mentoring, such as the importance of communication, consistency, and accessibility. In contrast, concrete strategies to improve the mentor–mentee relationship have been more difficult to define [6]. Funding agencies such as the National Science Foundation (NSF) and National Institutes of Health (NIH) have supported the implementation of policies aimed at improving this relationship. For example, in 2009, the NSF began requiring the inclusion of mentoring and development plans in grant proposals that request support for postdoctoral fellows; similarly, in 2014, the NIH announced (NOT-OD-14-113) that annual grant progress reports would be required to describe whether and how individual development plans (IDPs) are used to manage the career development of predoctoral and postdoctoral trainees. While a development plan can be effective as a tool to help a mentor and mentee work towards the mentee's long-term goals, this document alone is insufficient as it does not address the day-to-day operations of the lab—the source of many conflicts for both mentors and mentees.

Another resource that provides guidance on developing positive mentoring relationships is the "Compact Between Postdoctoral Appointees and Their Mentors" released by the Association of American Medical Colleges (AAMC) [7, 8]. Included in this compact are general principles governing responsibility for career development, development of the research plan, the need for regular feedback, and ethical conduct. However, this society-level document is lacking in details that would outline how these specific guidelines will be followed. Therefore, we have found it effective to develop "expectations documents"—lab-specific documents that detail both big picture elements of the mentor–mentee relationship as well as some of the nitty-gritty rules of how the lab operates. By clarifying the norms for a particular lab, an expectations document can provide a mechanism for prospective mentees to evaluate if a lab will be a good match for their needs. So, how do you develop your own expectations document? See our 10 simple rules below!

## Rule 1: Write it down

This may seem obvious, but it is important to remember that a written document more clearly and consistently communicates your expectations than conversations. Written documents also allow for both mentor and mentee to revisit the expectations as the mentoring relationship develops and provide documentation should a situation arise where either mentor or mentee does not adhere to the predetermined expectations. While it can be challenging to construct the initial document, even an incomplete draft could offset major clashes between mentor and mentee. To get started, you may be able to get a template from your institution, department, or training program, which will incorporate university policies and procedures. Alternatively, to help you as you start this process, we have included a sample expectations document (Suppl. File 1). This document is a modified version of the documents we utilize with our graduate student trainees. Which brings us to Rule 2.

## Rule 2: Tailor the expectations document to your audience and environment

In most research labs, there are personnel at a variety of career stages—postdocs, graduate students, undergraduates, and other scientific staff. Each of these groups has unique needs to address; as a result, it is useful to have separate documents for different personnel groups. Examples of graduate specific elements in the provided sample include indicating that the student is responsible for fulfilling course requirements, but that the mentor is available to help guide these decisions. For undergraduates, you may choose to discuss your grading policy, while for staff you may discuss their role in lab management, and the version for postdoctoral researchers may emphasize expectations regarding leadership and independence. You will also want to tailor your document to the type of research environment that your mentees work in. As discussed in a later rule, expectations may differ for research settings that are theoretical, computational, experimental, fieldwork based, or a combination of these environments.

## Rule 3: Convey the big picture

Ideally, the expectations document should provide the mentee with an understanding of your lab culture and approach to their training. Providing an overview of the lab environment as well as describing your mentoring philosophy can assist the mentee in establishing a positive relationship with both you and the other lab personnel. This information can also help prospective mentees determine whether your lab is an environment where they can picture themselves thriving. In our example, we provide both an overview paragraph summarizing these elements as well as comments throughout that relate our mentoring philosophy.

## Rule 4: But don't forget the nitty gritty

At this point, you may be wondering if it would be easier to use the published mentoring guidelines from the AAMC [7, 8]. While these guidelines provide an excellent source for developing your big picture philosophy, in our experience it has been beneficial to move beyond the mentoring philosophy and also convey some of the specific rules of the lab. It is not feasible to concisely list all guidelines related to lab performance or work expectations—however, clearly stating these rules can prevent significant conflict in the mentoring relationship. In our example document, we discuss hours and vacation, detail the overall requirements for lab safety and lab jobs (leaving further specifics to our lab protocols), conflict resolution, and outline how authorship is determined. For research that is theoretical and/or computational, it may be important to discuss policies on working remotely and documentation requirements for codes, while for fieldwork, discussion of expectations related to conduct and safety would be appropriate. Ultimately, you will want to confirm that the expectations that you outline for your mentees are consistent with the rules and regulations of your institution.

## Rule 5: Expectations are a two-way street

Just as you will outline your expectations for the mentee's behavior, it is important to outline what they can expect from you. Mentoring styles differ, and alignment between mentoring style and a mentee's self-identified needs can benefit both parties. For example, a student who wants regular feedback may struggle while working with a mentor who prefers a hands-off approach.

## Rule 6: Articulate boundaries

When constructing your expectations, be mindful of the power differential that exists between you and your mentee. The expectations document may be used to communicate professional boundaries, such as whether the mentee will be expected to contribute to work commonly performed by the mentor (e.g., our example includes discussion of assistance with grant preparation, advising other group members). Additionally, you can use the expectations document to articulate personal boundaries. For example, to maintain work–life balance, we have included information in our example on how much time a student should expect for answers to their questions and situations where it would be appropriate for the mentee to call on a personal number.

## Rule 7: Work with others to develop your expectations document

Are you feeling stuck or overwhelmed? Getting input from people with different perspectives may make it easier for you to develop your expectations document and determine sections that need more detail or clarification. For example, you may want to discuss your document with your own mentors, colleagues, or your more senior mentees. One especially effective strategy is to develop a small writing group with a few colleagues where each member develops an expectations document over the span of several meetings. In addition to their insights, the peer pressure to have a completed document for the next meeting may help to motivate you to complete this task.

## Rule 8: Plagiarism is okay (sort of)

As you look through examples or work with your colleagues, it is likely you will find statements that resonate with your approach. Because one aim of drafting an expectations document is to simplify the job of being a mentor, we would encourage you to ask for permission to copy and/or modify existing statements. Consistent with this, we grant permission for you to copy and/or modify sections of the example expectations document (Suppl. File 1). However, we encourage you to think critically and be certain that any statement that you use truly reflects your actual mentoring approaches—this is essential to prevent sending mixed messages to your mentees.

## Rule 9: Encourage regular conversation about the expectations document's interpretation

When first starting in a lab, a mentee's understanding of the expectations document will be largely theoretical. However, as the mentee progresses through their training and sees the mentoring expectations put into practice, new questions may arise regarding the interpretation and implementation of these guidelines. Regular conversations about the expectations document can help maintain an open channel of communication, head off misunderstandings, and provide feedback for document revision. In addition to informal conversations, it may be beneficial to set aside a part of one group meeting each year for this or incorporate it into your lab's evaluation process. These conversations lead to our final rule.

## Rule 10: This is a living document

As noted in Rule 1, it is appropriate to start with a smaller expectations document and add or refine content over time as needed. Even for those who start with a complete expectations document, unforeseen situations will arise. In addition, the rules of the graduate program or institution may change over time. Regular revisions to the expectations document allow for these changes in expectations to be incorporated so that all members of the lab remain on the same page.

## Conclusions

Like any other type of relationship in a person's life, the relationship between a mentor and a mentee requires intentional effort and clear communication to be healthy and successful. Providing your mentees with a guiding document about the expectations in your research lab benefits all parties. Sharing this expectations document with prospective lab employees can help them assess whether your lab is an environment that is likely to meet their needs and help you avoid a hiring mismatch. Once a mentee has joined your lab, the presence of written expectations can reduce the potential for conflicts and misunderstandings, which are damaging to the productivity and happiness of both the mentor and mentee. We hope that these 10 simple rules help you to develop an expectations document that works for your lab in order to lessen conflict and improve productivity.

## Supporting information

**S1 File. A sample mentor–mentee expectations document.**
(DOCX)

## Acknowledgments

We gratefully acknowledge Dr. Ben Cosgrove for suggesting we write up our experiences with expectations documents, Dr. Christine Pfund for her helpful discussions, and our mentors and mentees who have helped to shape our individual mentoring styles.

## References

1. Dolan E, Johnson D. Toward a holistic view of undergraduate research experiences: an exploratory study of impact on Graduate/Postdoctoral mentors. J Sci Educ Technol. 2009; 18(6):541–9.
2. Solorzano D. The road to the doctorate for California's Chicanas and Chicanos: a study of Ford Foundation Minority Fellows. Berkeley: California Policy Seminar; 1993.
3. Making the Right Moves: A Practical Guid to Scientific Management for Postdocs and New Faculty. 2nd ed: Burroughs Wellcome Fund and Howard Hughes Medical Institute; 2006.
4. Pfund C, Branchaw J, Handelsman J. Entering Mentoring. 2nd ed: W.H. Freeman; 2015.
5. Barker K. At the Helm: A Laboratory Navigator: Cold Spring Harbor Laboratory Press; 2002.
6. Pfund C, Byars-Winston A, Branchaw J, Hurtado S, Eagan K. Defining Attributes and Metrics of Effective Research Mentoring Relationships. AIDS Behav. 2016; 20 Suppl 2:238–48.
7. Compact Between Biomedical Graduate Students and Their Research Advisors. Association of American Medical Colleges; 2017.
8. Compact Between Postdoctoral Appointees and Their Mentors. Association of American Medical Colleges; 2017.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for successfully completing a graduate degree in Latin America

**Ernesto Ruelas Inzunza[1]*, Gabriela I. Salazar-Rivera[1☯], Magdiel Láinez[1☯], María Guadalupe Ruiz-Gómez[1☯], Carlo A. Domínguez-Eusebio[1☯], Griselda Cristóbal-Sánchez[2☯], Issaac A. Teodosio Faustino[3☯], Edel Pérez-López[4‡], Meagan L. Campbell[1‡], Marcus Vinicius Merfa[4‡], Ivonne Tatiana Latorre Beltrán[5‡], Fernanda Armas[6‡], Claudio Mota-Vargas[6‡]**

**1** Instituto de Biotecnología y Ecología Aplicada, Universidad Veracruzana, Xalapa, Veracruz, Mexico, **2** Universidad Veracruzana, Centro de Investigaciones Tropicales, Xalapa, Veracruz, Mexico, **3** Facultad de Biología, Universidad Veracruzana, Xalapa, Veracruz, Mexico, **4** Department of Entomology and Plant Pathology, Auburn University, Auburn, Alabama, United States, **5** Independent researcher, Bogotá, Colombia, **6** Instituto de Ecología A.C., Xalapa, Veracruz, Mexico

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* ruelas.uv@gmail.com

Check for updates

OPEN ACCESS

**Citation:** Ruelas Inzunza E, Salazar-Rivera GI, Láinez M, Ruiz-Gómez MG, Domínguez-Eusebio CA, Cristóbal-Sánchez G, et al. (2017) Ten simple rules for successfully completing a graduate degree in Latin America. PLoS Comput Biol 13(9): e1005682. https://doi.org/10.1371/journal.pcbi.1005682

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** September 21, 2017



**Funding:** Mexico's Consejo Nacional de Ciencia y Tecnología provided graduate fellowships for most of the student authors of this paper and fellowship 47135 of the Sistema Nacional de Investigadores for ERI. EPL and MVM are funded by an Agriculture and Food Research Initiative competitive grant 2016-70016-24844 (USDA National Institute of Food and Agriculture, Citrus Disease Research and Extension). ERI also thanks Mexico's PRODEP Program of the Secretaría de Educación Pública. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

The successful completion of a graduate program in biological sciences is an endeavor that entails a passion for science. It requires some essential ingredients, such as the ability to think independently, a deep understanding of your study system, focus on strategic priorities, and establishing collaborations. A graduate program is also filled with challenges and hurdles. These and other key items have been covered in the literature, including papers in the "Ten Simple Rules" series [1, 2].

We have found, however, that in addition to those commonalities, students face regional variations, features that are specific to a particular part of the world. In this article, we posit that certain circumstances faced by graduate students are distinctively Latin American. On one side, limitations in human, infrastructure, and monetary resources are features that underlie the performance of the majority of graduate programs around the developing world [3]. On the other side, many unexplored ideas and research avenues dazzle the future possibilities of students of this region. Regardless, there is something unique to Latin American programs that needs to be explicitly said. We hope that the following 10 rules can appropriately summarize those characteristics.

The aim of this paper is to help prepare students interested in graduate programs in Latin America for the exciting experience they are about to start. When reading each of these rules, keep in mind the goal of this paper is to power your interest in science by presenting you its brilliant side as well as its caveats. We put special emphasis on how creativity, collaborations, and a great dose of energy can help them overcome those limitations.

## Rule 1: Investigate what you are getting into

Choose your adviser and your program carefully [1]. Graduate programs tend to have a local flavor and specialization. Consequently, many of them lack a critical mass in some areas. It is difficult to find a world expert in your specific field that—in addition to having similar research interests as yours—can provide funding or materials to support at least part of your work. Students often include a codirector for their graduate work, who usually brings

**Competing interests:** The authors have declared that no competing interests exist.

additional expertise and other resources to their research but that often puts them in the midst of opposing views of how their own work should be done. However, 2 experienced views and 2 sources of financial support (even if meager) are better than 1.

If at all possible, meet your future adviser(s) in person. This is of vital importance. Culturally, these meetings are more valued than other forms of communication. If you don't find an adviser with closely matching interests, a codirector is often the way to make 2 research topics meet.

Students entering a master's degree normally initiate their program with a broader set of interests than students entering a doctoral program (for whom a closer focus is expected). In both cases, the refinements of their investigations will come with time and in the process of crafting a research project. Be open-minded and concede on some of your initial views of work, but do not get into doing something you don't want to do. Enter a field you feel passionate about.

Figure out the Latin American cultural idiosyncrasies regarding students; learn about the habits of your adviser(s), how much time and attention they give to them, how willing they are to go out of their way to help their students, and how good they are at communicating. Generally speaking, senior, better-funded scientists have less time for students.

References of the specific program you are applying to are key. With very few standardized, region-wide, graduate program rankings (e.g., Brazil's Fundação CAPES), a way you can gain insights into the program you are interested in is by researching its faculty, department, and institution using Google Scholar, ResearchGate, and Academia.edu. Word of mouth matters, especially in Latin America. Ask current or past students of the program about their experiences. Beware of (often private university–offered) "duckie" degrees. The basic infrastructure of an institution is a criterion of utmost importance, particularly if your research is demanding in terms of equipment, space, computing, and laboratory conditions. Consider having international committee members (often a great resource [4]).

For those interested in professional futures outside academia, there are very few programs preparing students for careers in private, non-profit, and government positions. The majority of the opportunities for growth in those sectors are accessed through your own enthusiasm, life experience, and hard work.

## Rule 2: Financial resources for graduate students are awarded on a competitive basis

Some countries have robust scholarship programs for graduate students, often open to applicants from any country (provided they pass through the regular admissions procedures). Funding is usually limited or highly competitive. Take advantage of these programs, as they are constantly adjusting the number of scholarships available and the amount of financial support that they provide. Some graduate programs (very few, we should say) have their own additional funding to supplement graduate students. Most scholarship programs are mutually exclusive—accepting one precludes you from applying to a second one. Working on other jobs while you are a graduate student, when allowed, is discouraged and often limited to a certain number of hours per week. In some countries, scholarships are actually student loans (the so-called "credit scholarships").

Consider that the standard monthly scholarship is only sufficient to cover the expenses of a single person and that having dependents only makes you eligible for additional funds under some circumstances. Many students report using some of their scholarship funds to cover parts of their research expenses.

Learn how to get funds from external sources (Box 1). Go beyond the national boundaries for this; small grants are available from international, and even from nonconventional (e.g., private), sources for graduate students. Being a student member of scientific societies makes you eligible for larger pool of grants.

## Box 1. How to find funding for graduate student research in Latin America

Looking for funding to cover your graduate research can take you to an online labyrinth. Below, we list 4 common gateways to find what you are looking for.

**Graduate fellowships.** The funding to cover the bulk of your graduate studies typically comes from fellowships offered by federal research agencies. Examples of in-country fellowship programs for graduate students are Mexico's Consejo Nacional de Ciencia y Tecnología (http://conacyt.gob.mx/index.php/becas-y-posgrados), which offers fellowships for graduate studies and postdoctoral research in Mexico. These fellowships are available to applicants from any country or for Mexican nationals who plan to study abroad. In Argentina, the Consejo Nacional de Investigaciones Científicas y Técnicas (http://convocatorias.conicet.gov.ar/becas/) works similarly, with additional grants in support for research projects from graduate students. In Brazil, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (http://www.cnpq.br/web/guest/apresentacao13/) supports graduate work. Many of these agencies have special opportunities for women in science, special projects, internationalization, and study abroad opportunities that are separate from their standard graduate fellowships.

**Scientific societies.** The majority of the most prominent scientific societies in the world have lists of grants and prizes available for students. Are you an ecologist? Check the website of the Ecological Society of America (https://www.esa.org/esa/careers-and-certification/funding-and-grants/). Perhaps an evolutionary biologist? Find the grants section of the Society for the Study of Evolution (http://www.evolutionsociety.org/). For earth and space scientists, the American Geophysical Union has funding listings as well (http://education.agu.org/grants/). Biochemists and molecular biologists can find funding opportunities in the American Society for Biochemistry and Molecular Biology website (https://www.asbmb.org/careers/scholarshipsandawards/). This list is a tiny sample. Check the society that most closely matches your field and take advantage of this benefit to its members.

**Courses, workshops, and conferences.** The organizers of training and academic exchange gatherings, such as conferences, symposia, and its associated courses and workshops, normally have funding available to facilitate graduate student attendance. This support has many different forms: registration fee waivers, free society memberships, volunteering opportunities, and competitive travel grants for attendants. Explore your possibilities carefully and put time into applying for relevant opportunities.

**Grants and prizes from subject-specific sources.** Last, many groups ranging from scientific to amateur make funding available for graduate students through small grant programs. Google them using the search words that best fit your research. Here is the top hit of the search string "grants + bird + research": http://ornithologyexchange.org/resources/grants.html. It is a list of 246 sources of funding ranging from a few hundred dollars to several hundred thousand (with a median of about $4,000). Latin American students are eligible to many of them.

Some students entering graduate programs that already have academic positions can take a salaried leave of absence. If you are going back to grad school from a job elsewhere, remember that becoming a student is a very demanding commitment and, consequently, comes with changes in your income. Save money. Project future expenses. Be ready for periods where your savings will be your best bet. Graduate school is a full-time endeavor.

## Rule 3: Master administrative procedures from admission to graduation

Contrary to the somewhat standardized admissions processes at universities in the developed world, the process to select graduate students varies greatly in Latin American programs (often within the same university). Deadlines, requirements, exams, interviews, etc. are rarely the same across programs of the same discipline. This means that each and every application requires a substantial effort to put together.

Follow admission rules strictly. Be aware that some graduate programs include prep courses for their own degrees. Standard exams are often the most valued criterion for selecting students and a filter to reduce the size of the group of applicants.

Throughout the duration of your program, be meticulous when facing administrative procedures. Latin American graduate programs are plagued by paperwork requirements. The systems themselves can be rigid and not prone to creative solutions. When the specific requirements of a program do not match the paperwork that exists in your country, the first answer is frequently "no." Be persistent and try to overcome that first no. Make in-person visits to program officers, get to know the registrar and administrators, be strict and exhaustive when compiling certificates, apostilles, notarized documents, immigration paperwork, and other forms, letters, and signatures. Your management of those procedures and requirements needs to be spotless. Plan way ahead of time and always meet deadlines.

Officers running graduate programs often do many jobs at once. We know programs run by people who have a research program of their own, teach, and coordinate the graduate program at a university or research center. Be persistent when communicating with them. Emails very often do not suffice. Insist. Get on the phone. Get answers to your questions. And do so with grace and good manners.

## Rule 4: The English language is essential

Let us say it again: It is essential [4, 5]. The overwhelming majority of your reading materials while you are a graduate student are in English. The language to communicate your results in peer-reviewed journals and reach broad audiences is English as well. The standard language used in international conferences and symposia is English. Many courses, software, listservs, newsletters, scientific societies, and a myriad of other tools to catapult your science career are, as you can imagine, in English. You will need to work hard in overcoming the limitation of English not being your native language. You won't go very far without it. Better embrace it, learn English, and do it sooner rather than later.

With the English language as one of your skills, you should be able to access and navigate the somewhat flat field of science. The cost of not knowing English is too expensive, and you cannot afford it.

## Rule 5: Learn how to write

Papers are the most important communication currency in the sciences. Understand their elements: the main narrative, tables, figures, appendices, and other ancillary/supplementary information. Journals have adopted a variety of formats for the publication of research reports, opinion essays, letters, and other types of papers.

Many graduate programs have courses and training opportunities to learn how to write papers. These courses teach how to build a manuscript around a central idea, organize its content, write clear and concise messages, and be impeccable with regards to grammar, spelling, and punctuation. They also teach what, where, when, and how to publish.

Learning good writing is a complex process. It is another essential skill you must have to succeed as a scientist.

## Rule 6: Economic and logistical limitations are key criteria for selecting study subjects

Field site safety, social unrest, logistical complexity, distance to work sites, equipment availability, and qualified technical expertise are examples of central aspects that need to be carefully gauged when embarking on a research project. Many of them can be overcome with enough energy and time spent writing grants and establishing collaborations, but know that some items may just be way beyond what is reasonable or doable.

## Rule 7: Time is money or "El tiempo es oro"

Stay focused and solve—as soon as you can and in the best way possible—all those personal commitments that could distract from your work in the upcoming years. Figure out all housing, transportation, and family issues that can be solved ahead of time. Seek help or put away from yourself the additional headache of having to deal with problems that could have been solved ahead of time while being a graduate student. Learn about how to manage time, financial resources, and people as part of your training. Always bear in mind the expected standard duration of a doctoral program is 4 years and just 2 for a master's program (Rule 1 in [2]).

It is true that family closeness and tightly knit social groups are much stronger in Latin America than in other regions of the world (and can be an amazing help), but they can also be a time, effort, and emotional distraction from your course of study. Do your best to find a balance of their costs and benefits. Many graduate students can overcome the heavy weight of being the pillar of the family, but only with Herculean efforts that almost invariably affect the quality and quantity of their graduate work.

## Rule 8: Make the best out of the courses you have at hand

Graduate programs in Latin America can be much broader in scope than you wish. Many of the programs, by necessity, have "fixed" courses that every student has to take and electives that aren't quite as diverse and specialized as you need. Many programs have limitations on the number of courses you can take at other institutions (and issues over the compatibility of the evaluation point system arise quickly). Graduate programs often recognize such limitations and encourage you to exploit the collaboration connections of your university (some of them international [4]). Put extra time and effort into finding high-quality courses, even if they take you to faraway places. Find out if your university has ongoing formal agreements with other programs and make sure coursework offered elsewhere is compatible with your program. Find out if you need or can revalidate coursework taken at other places. Complete your coursework as soon as possible (you will need that time and energy in order to develop a strong research project).

## Rule 9: Give your own skills and potentials the right dimension

Be sure you are fair in dimensioning your own strengths and potentials. Become a true expert in your field! Know your stuff: Speak and write about it firmly, but don't be cocky. Know that graduate students endure similar problems anywhere in the world; e.g., did you ever read about imposter syndrome [3] or writer's block [5]? Low self-esteem, a common problem faced by students while in graduate school, seems to be particularly acute in Latin America.

Do not wait for your adviser to be the source of all initiative. Take the right position in the "autonomy/authority" continuum (Fig 1, Box 2). Follow Rule 2 in [1] closely. Being a graduate student is not a temporary job for your résumé; it is the preparation period for your career in science.

## Rule 10: Be proactive, creative, collaborative, and self-taught

You have in your hands the ability to clear obstacles you may find during your graduate program. Collaborate with other students and research scientists. Take advantage of the unprecedented opportunities to communicate, travel, take online courses, and learn from others in active exchanges via email both nationally and internationally [6]. Take a year abroad (it is a strong addition to your CV). Be open to criticism from your peers and professors. Ask about what you don't know (the worst question is the one that is never asked). Circulate your proposals, papers, and ideas for comments and reciprocate the help that you have received from others.

Join scientific societies. Be active in their committees, attend their meetings, and volunteer at their conferences and symposia; present your research data there, get feedback, and network with other students and professionals. These experiences are crucial in your professional development.

In most Latin American graduate programs, obtaining teaching skills is not part of your graduate training. If you are seeking a career in academia, make sure you find a way to get this experience.

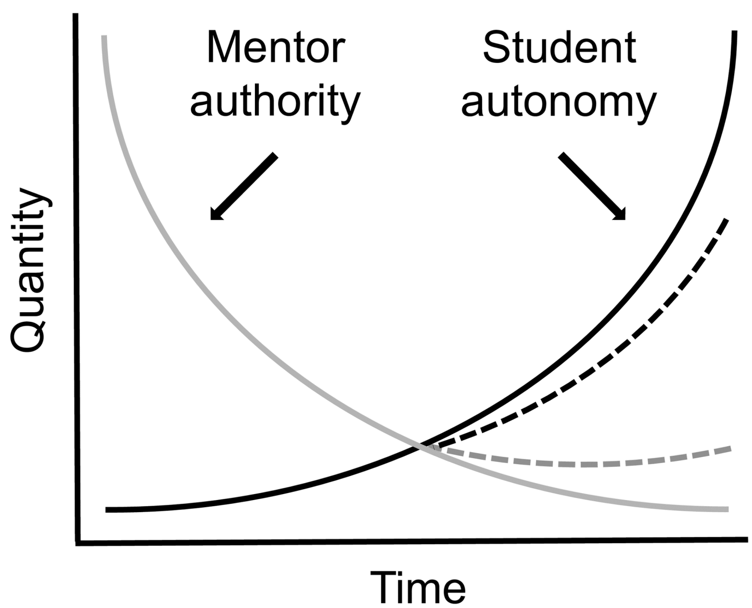



**Fig 1. An equilibrium model of academic authority and autonomy.**

https://doi.org/10.1371/journal.pcbi.1005682.g001

### Box 2. Academic autonomy and authority in graduate programs

In graduate education, the academic authority of teachers and advisers decreases—or should decrease—as a function of time (Fig 1). In the figure above, this reduction in the role of mentors (illustrated as a gray curve) goes down as the autonomy of students (black curve) increases. The degree to which advisers concede such a gain in autonomy varies depending on the mentoring style, the strength of student skills, and the culture of the professor–student relationship. Conversely, students are able to attain autonomy of their graduate research depending on how strongly they feel about their own skills. These differences are illustrated in the model above by dotted lines that represent the variance of autonomy-authority curves. Graduate students are expected to gain a lot (in their master's degrees) to almost absolute (in their doctoral degrees) autonomy throughout the course of their studies. One of the aspirations of advisers is that their advisees have gained substantial expertise in their field by the time of degree completion.

Learn the basic tools of the trade in your discipline: equipment, analysis techniques, software, etc. Engage, on your own or in groups, in self-study and in journal/book clubs.

## A corollary

Latin American graduate programs have made great strides in producing excellent graduates and future scientists. Now you can find internationally outstanding and highly productive research groups in many countries. Many foreign students come to Latin American graduate programs to enjoy a rich cultural experience.

Perhaps the greatest reward of completing a graduate degree in Latin America is that many (most?) of the study systems haven't been as thoroughly studied as in the developed world. Your research can make giant discoveries and provide the very first insights into a particular question or be the benchmark work that finds the solution to a big problem. You can make a difference by becoming a finely trained mentor that guides the careers of many more young scientists and professionals, influencing policy, governance, and public opinion.

Make a commitment to pay back to the country and people that helped you to do what you've done as part of your graduate research. Besides publishing all of your research, those investment returns can take the form of actions to bridge your work with issues of public interest: papers for a general readership, manuals, catalogues, and conferences for a broader audience. Some of us think the social meaning of Latin American science should privilege applied (over basic) research and focus on the solution of society's most poignant problems (others do not, but that's okay: basic science, when properly conducted and published, will prove useful to mundane applications at some point in the future). Avoid academic endogamy. Consider changing to an entirely different program, institution, city, and/or country when you are ready to your next step (a PhD if you are a MSc student, or a postdoctoral position). Make the broadening of your horizons a priority of your training.

## Acknowledgments

The authors of this paper are from 5 different Spanish-, Portuguese-, and English-speaking countries (Mexico, the United States of America, Brazil, Cuba, and Colombia) and requested feedback from colleagues from Argentina, Peru, Ecuador, Costa Rica, and Chile. The authors have all studied, taught, or done research in graduate programs in Latin America. We thank the graduate programs of the Instituto de Biotecnología y Ecología Aplicada and the Centro de Investigaciones Tropicales of the Universidad Veracruzana for their support for our work. The authority-autonomy model in Box 2 is a result of discussions with T. M. Schusler of Loyola University in Chicago.

## References

1. Gu J, Bourne PE. Ten simple rules for graduate students. PLoS Comput Biol. 2007; 3(11): e229. https://doi.org/10.1371/journal.pcbi.0030229 PMID: 18052537
2. Marino J, Stefan MI, Blackford S. Ten Simple Rules for Finishing Your PhD. PLoS Comput Biol. 2014; 10(12): e1003954. https://doi.org/10.1371/journal.pcbi.1003954 PMID: 25474445
3. Clance PR, Imes SA. The imposter phenomenon in high achieving women: dynamics and therapeutic intervention. Psychotherapy: Theory, Research and Practice. 1978; 15(3): 241–247.
4. Moreno E, Gutiérrez J-M. Ten Simple Rules for Aspiring Scientists in a Low-Income Country. PLoS Comput Biol. 2008; 4(5): e1000024. https://doi.org/10.1371/journal.pcbi.1000024 PMID: 18437198
5. Oliver LJ Jr. Helping Students Overcome Writer's Block. J Reading. 1982; 26(2): 162–168.
6. de Grijs R. Ten Simple Rules for Establishing International Research Collaborations. PLoS Comput Biol. 2015; 11(10): e1004311. https://doi.org/10.1371/journal.pcbi.1004311 PMID: 26447799

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for surviving an interdisciplinary PhD

**Samuel Demharter[1☯]*, Nicholas Pearce[2☯], Kylie Beattie[2], Isabel Frost[1], Jinwoo Leem[2], Alistair Martin[3], Robert Oppenheimer[4], Cristian Regep[3], Tammo Rukat[2], Alexander Skates[2], Nicola Trendel[2], David J. Gavaghan[3], Charlotte M. Deane[2], Bernhard Knapp[2¤]***

**1** Doctoral Training Centre for Systems Biology, University of Oxford, Oxford, United Kingdom, **2** Doctoral Training Centre for Systems Approaches to Biomedical Science, University of Oxford, Oxford, United Kingdom, **3** Doctoral Training Centre for Life Sciences Interface, University of Oxford, Oxford, United Kingdom, **4** Doctoral Training Centre for Synthetic Biology, University of Oxford, Oxford, United Kingdom

☯ These authors contributed equally to this work.
¤ Current address: Faculty of Medicine and Health Sciences, International University of Catalonia, Barcelona, Spain
* samuel.demharter@dtc.ox.ac.uk (SD); bknapp@uic.es (BK)



OPEN ACCESS

**Citation:** Demharter S, Pearce N, Beattie K, Frost I, Leem J, Martin A, et al. (2017) Ten simple rules for surviving an interdisciplinary PhD. PLoS Comput Biol 13(5): e1005512. https://doi.org/10.1371/journal.pcbi.1005512

**Published:** May 25, 2017



**Funding:** The DTC at the University of Oxford is funded by the EPSRC, BBSRC, and MRC. Grant numbers for the individual programmes are as follows: Systems Biology (EP/G03706X/1), Systems Approaches to Biomedical Sciences (EP/L016044/1, EP/G037280/1), Life Sciences Interface (EP/F500394/1, EP/I017909/1), Biomedical Imaging (EP/L016052/1), Synthetic Biology (EP/L016494/1), Synthesis for Biology and Medicine (EP/L015838/1), Doctoral Training Partnership (BB/M011224/1). The funders had no role in the preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Many of today's pressing research challenges require a multifaceted approach that combines several historically distinct disciplines. As a result, there has been a surge in funding for interdisciplinary PhD programmes. Some examples include the United States National Science Foundation (NSF) Research Traineeship (NRT) [1] (succeeds the Integrative Graduate Education and Research Traineeship [IGERT] [2]); the European Research Council's Innovative Training Network (ITN) [3]; and, in the United Kingdom, strong growth in interdisciplinary doctoral programmes across all research councils, led by the Engineering and Physical Sciences Research Council (EPSRC) and their Centres of Doctoral Training (CDTs), with the strong support of UK universities and industrial partners [4].

First and foremost, an interdisciplinary PhD is a great chance for students to pursue truly novel research, a range of different career paths, and a stimulating intellectual life. However, these benefits are often accompanied by additional academic and logistical challenges. The rules presented here aim to provide guidelines that will enable PhD candidates to maximise the benefits of interdisciplinary research whilst minimising any burdens.

The term "interdisciplinary" itself has many different meanings in common usage; for the purposes of this article, we define "interdisciplinary" as the synthesis of 2 or more disciplines, establishing a new level of discourse and integration of knowledge [5]. Whilst most of the advice for students considering interdisciplinary programmes is similar to that of traditional graduate programmes [6], there are also differences that students should be aware of and prepare for; the interdisciplinary PhD programmes described above, and interdisciplinary research in general [7], bring unique opportunities as well as challenges.

There are several different types of interdisciplinary PhD programmes, and their organisation varies widely from country to country. In the UK, interdisciplinary programmes are increasingly funded through CDTs, as mentioned above; investigation of the major funding bodies in your field is therefore a good place to start for discovering which programmes are available. Different interdisciplinary PhD programmes may also be organised very differently: courses might be very structured, including preparation courses and short rotation projects before the PhD; some might include continuous training throughout the PhD; and some may require you to teach, while others may not. Since interdisciplinary programmes are by their nature highly varied, some also allow you to start the programme without first having chosen a supervisor; this enables you to familiarise yourself with different fields before choosing a PhD topic. For this type of programme, it is less important to have a particular subject area in mind. However, if you do, it is important to investigate the research groups associated with the program before you apply to ensure that the course is compatible with your aims.

We are a group of PhD students and programme directors at the Doctoral Training Centre (DTC), University of Oxford. The Oxford DTC was founded in 2002 and has since accommodated over 550 students across 7 interdisciplinary programmes, namely: EPSRC Life Sciences Interface, EPSRC Systems Biology, EPSRC and Medical Research Council (MRC) System Approaches to Biomedical Science, EPSRC and Biotechnology and Biological Sciences Research Council (BBSRC) Synthetic Biology, EPSRC Synthesis for Biology & Medicine, EPSRC and MRC Biomedical Imaging, and BBSRC Doctoral Training Partnership. Drawing from our collective experience at the DTC, we present 10 simple rules for surviving and thriving in an interdisciplinary PhD. We highlight the importance of having a comprehensive plan with a realistic time line, maintaining good communication between all supervisors and collaborators, and making the most of the intellectual freedom that you are provided with when working in an interdisciplinary setting.

## Rule 1: Involve everyone in the planning, and make contingency plans

An interdisciplinary PhD will most likely require acquiring a wide range of new skills, involve more than one supervisor, and depend on multiple collaborators. It is therefore imperative to include all your supervisors early on when discussing and deciding the goals and aims of the PhD. Likewise, you should collectively lay out the plan to achieve those goals. This way, the expectations and requirements of everyone are more effectively managed (see Rule 2).

As with all PhDs, it is beneficial to generate a time line with specific milestones. This will allow you to assess the progress you have made throughout your PhD and identify any potential problems. Keep in mind that the PhD plan will change with time, so allow yourself room to manoeuvre. Importantly, make sure you have a contingency plan that covers eventualities such as collaborators not delivering data or dropping out entirely.

When planning your project, try to look ahead in terms of getting the right mentors and support network (see Rule 5). Ask your supervisors to make introductions; this can make it much easier to meet people, especially at the beginning of the PhD when you are still finding your feet.

Finally, ensure that you understand your administrative obligations. Depending on your institution and department, there may be different requirements for each stage of progression in a PhD. While the high degree of flexibility may be beneficial and allow you to choose the requirements that suit you best, it also means that you are more likely to fall outside of the normal procedures. As a result, the administration is not always prepared for the requirements of interdisciplinary students. It is therefore, as always, important to try to plan your last year in advance [8]. This will reduce the number of times you run into unexpected hurdles and make the final months less daunting. Take initiative and think ahead.

## Rule 2: Be a diplomat: Start managing expectations early on

Try to organise regular meetings with all your supervisors and collaborators. As well as ensuring progression of the PhD, collective planning and discussion helps to prevent frustrating situations and disagreements. Recognise that supervisors from different fields may have very different expectations for what is achievable in a particular time frame or may find it hard to judge the difficulty and likely time span of research outside their own area of expertise.

If questions regarding the direction of the PhD arise, such as the best approach to a problem, it is normally best to discuss the problem directly with all participating parties in the same room.

With several parties involved in your PhD, it is essential to keep communicating on a regular basis, with regular time slots for video conferences or face-to-face meetings.

## Rule 3: Define the boundaries of your research: Explore and familiarise, then be pragmatic

Once everyone is on the same page and you have laid out a plan for your PhD, you need to start to do your research. A "traditional" PhD student quickly develops very deep knowledge of a narrow subject area in a particular discipline, whereas an interdisciplinary student is likely to obtain knowledge that is less deep but spread out across several subject areas and multiple disciplines. It is therefore important to anticipate and explore the fields relevant to your PhD early on. Attending a wide range of seminars or even undergraduate lectures is a good way to gain a foundation of understanding in a new field and to learn discipline-specific terminology. Time spent investigating these complementary subject areas early will be beneficial in the long run, as it will enable you to see the bigger picture and place your work in context.

Furthermore, the quality, quantity, and structure of data vary between disciplines [9]. Make sure you know what to expect, and perform "sanity checks" on the data before you use it for anything. This will enable you to identify any issues and allow you to take the necessary steps to interpret the data correctly [10].

Whilst it is important to become familiar with all the relevant topics across the disciplines of your work, you cannot learn everything. As the PhD progresses, start focusing on the core areas that are directly relevant to your research. Consider a funnel-shaped learning approach, where you learn the fundamentals of the field first and narrow down on more specialised topics at later stages. Explore and familiarise yourself, but then try to be pragmatic and goal oriented.

## Rule 4: Don't be embarrassed: Always ask the "stupid question"

When you are exploring new fields, it is normal to feel estranged, alone, and lost. Seminars, group meetings, and research talks in this alien field can seem like they are being conducted in an entirely different language. It is not uncommon to attend meetings where unfamiliar jargon is heavily used. Consider immersing yourself and "spending time with the natives" as much as possible; this is often the quickest and most effective way to become proficient in the new field.

Being in between disciplines naturally means that a lot of time will be spent being a novice, and progress may feel slow at the beginning. You will always have to ask a lot of questions, and it is fine to solicit help shamelessly. You might feel like your questions are too simple to waste anyone's time with, but keep in mind that most people are happy to see you engaging with their subject. Furthermore, questions from people with a different background often lead to a new perspective that might not have been considered otherwise.

Do not be afraid to ask the "stupid" question; what seems trivial might not be quite so simple.

## Rule 5: Build a network: Find other people to complement the gaps

Due to the nature of interdisciplinary research, there will always be significant gaps in your knowledge. Identify fellow researchers early on in your PhD who complement your knowledge base; you will then be able to call upon each other's expertise when required. You might not find all the help you need in your department. A quick internet search and an email may help you to find the right group; most scientists are friendly and happy to share their knowledge. If in doubt, your supervisors should be able to put you in contact with the relevant people.

Fostering and maintaining relationships with researchers from diverse backgrounds is a key aspect of doing an interdisciplinary PhD; it will make your life a lot easier when it comes to finding someone to answer your questions, finding suitable collaborations, or getting your hands on that crucial data set. Try to find people with whom you enjoy working and who are good at what they do. It is worth spending some time on developing good relationships, as there is a good chance you will keep working together beyond the PhD. A paper specifically focusing on different aspects of cross-disciplinary collaborations is available in the 10 simple rules series [10].

When communicating with researchers in different disciplines, be sure to clarify what people mean by the terminology they use, as the same word may mean different things in different fields. As an example, consider the word "orthogonal": in geometry, 2 lines are orthogonal if they form a right angle; in statistics, independent variables that affect a dependent variable are considered to be orthogonal if they are uncorrelated; in taxonomy, a classification is orthogonal if each item is a member of only 1 group; and in biochemistry, the 2 types of DNA base pairs are considered orthogonal interactions. Ask for and give explanations for technical jargon, as the language barrier is always present in interdisciplinary collaborations.

While it is good to have some close contacts, it is also valuable to develop a more diverse and loosely connected network. You might consider this a "dormant" network that you can dip into when the opportunity or the need arises; it may just lead to a fruitful collaboration.

Many of the most useful conversations happen when you least expect it. Try to socialise with your peers, especially in nonacademic settings.

## Rule 6: Embrace your unique skillset and use it to redefine discipline boundaries

The value of symbiotic relationships between disciplines is well recognised. It comes from the realisation that some research questions are neglected because they do not fit inside the traditional boundaries of the discipline. Similarly, suitable methods or techniques for a particular problem might already exist in another field but have gone undetected due to the isolation of disciplines. As an interdisciplinary researcher, you are well placed to tackle those neglected problems. By embracing your unique skillset and looking for opportunities to connect the dots, you may find yourself redefining traditional boundaries or facilitating a groundbreaking translation of technology from one discipline into another.

This also holds true in collaborations. Often, solutions to problems that you would consider trivial are of high value to scientists in other fields. Recognise these opportunities and be open about your competencies, even the simplest ones. You may be surprised how much you can contribute. This is a good way to gain visibility among your colleagues and open new opportunities.

However, be careful, as it is easy to become a service provider to others in a different field when you have skills that they do not. You may not want to become the IT administrator responsible for routine data processing and organisation or the lab assistant responsible for routine and tedious experiments. Only consider doing these things if they have clear limits and clear benefits.

## Rule 7: Feel free to swim against the flow, to experiment, and to fail

Established fields will commonly have a dogmatic approach to certain problems that has evolved over generations of researchers. These well-established methods have been proven to work and have a vast amount of research to back them up. However, as a researcher with a potentially different view of the problems, you may be attracted towards approaches that are considered unconventional. Do not be afraid to challenge the dogma of the field, provided your approach has previously unanticipated benefits.

Furthermore, as someone who sits between disciplines, the techniques you require may not have been developed yet or may simply never have been applied to an individual discipline before. Developing new techniques (or applying "foreign" techniques) and proving their utility comes with higher risks and potentially higher rewards. As the forerunner, you will have to figure things out for yourself. You will get stuck and your approach may fail. Do not lose motivation during these seemingly unproductive phases. You may even encounter resistance to breaking dogmas from your supervisors, so it is important to remember that you can afford to spend time on satisfying your curiosity. Exploration involves failure, but it is fundamental and necessary in science, as it results in new discoveries and ideas.

## Rule 8: Plan your career and publish accordingly

By the end of the PhD, you will have a wide set of transferrable skills. These allow you to pursue careers in niche areas as well as more general fields. Furthermore, if you identify a skill that you do not yet have, an interdisciplinary PhD is a good opportunity to build the required competencies—so start thinking early about where you want to end up. Note that there will be different requirements depending on which career path you choose to follow. There are many places your skills could take you.

Another important factor in your career will be your publication record. Different career paths may require publishing in different journals, so you need to choose carefully. Working in an interdisciplinary setting means that you have a much larger variety of journals to choose from than a single-subject PhD, where there are usually only a select few journals.

Make sure you familiarise yourself with all relevant journals. This is especially important when you publish in journals with which your main supervisor has no experience. To narrow down the search, you can discuss options with colleagues in your field(s) and choose from a number of web services and journal selectors that find suitable journals based on information from your title, abstract, and other key words. The natural place to publish will most likely be the journals whose papers you most frequently read, but also look at the journals where these papers are frequently cited. Once you have identified some candidates, consider if the aims, scope, readership, and publication history of the journal are in alignment with your manuscript. Also, familiarise yourself with the rules of the journal by reading the author guidelines, and try to identify its readership and access options. When deciding on a journal, you may not feel that it is your choice, but try to engage in the discussion with your supervisors and explain your reasoning. An interdisciplinary PhD provides many career opportunities, as your unique skill set may be appreciated in a number of different sectors. Try to shape your PhD to your career plans and aim to publish accordingly.

## Rule 9: Adjust to your audience

Interdisciplinary researchers will have to present work at meetings, seminars, and conferences with vastly different foci (e.g., to both theoretical and experimental audiences). Your thesis will probably be examined by experts from different fields. Thus, you will have to adjust your language appropriately; your audience may comprise specialists in a particular field, those with mixed backgrounds, or those with no scientific training at all. This is, of course, applicable to all researchers, but the frequency of having to change a presentation to suit an audience is increased in interdisciplinary areas. Also, remember that you may have to run your proposed presentation past multiple supervisors and collaborators, so make sure to allocate enough time for this.

For a group of mixed backgrounds or a group with a different background, it is essential that you learn to anticipate any gaps in people's knowledge and provide a coherent story no matter how basic you think the material is. It may also be a good idea to explain concepts from several angles to accommodate as many people as possible. When describing an abstract or technical concept, avoid jargon and use visual representations where possible. Imagine explaining it to your nonexpert self before you started doing your research.

When asked to participate in science communication and public engagement events, another approach is required. A bird's-eye view, interesting examples, and clear language are crucial for effective communication. Do not underestimate the work that goes into these presentations; you will have to think about your research from a completely different perspective to your own. When in doubt, always start explaining things in the simplest terms possible and elaborate later on, if necessary. For high-profile events or key interview presentations, always practice in front of a friendly audience and act on feedback.

## Rule 10: Relax and enjoy

A PhD is a marathon, not a sprint. Make sure you are comfortable with your project and that you are in a position to enjoy the experience. The best publication record will be of little benefit if you do not enjoy the process. Fortunately, an interdisciplinary PhD often provides unique opportunities for you to design your research around your interests. Use this flexibility and shape your work into something that you enjoy and fully embrace.

For many scientists, creativity and productivity are highest during the first days after a break. Make use of your supervisors' and the university's resources, engage in clubs and other activities outside of your research, and take vacations when you need them. Taking time off will benefit you in the long run, as you will return to your research with renewed energy and a fresh mind.

A PhD is a huge investment of your time, energy, and creativity—the finale of years of education and training. By following these 10 simple rules, it should be a rewarding and empowering experience for you!

## References


1. National Science Foundation Research Traineeship (NRT) Program—Program Solicitation. NSF; 2015 Oct. Report No.: nsf16503.
2. Jennifer Carney, Alina Martinez, John Dreier, Kristen Neishi, Amanda Parsad. Evaluation of the National Science Foundation's Integrative Graduate Education and Research Traineeship Program (IGERT): Follow-up Study of IGERT Graduates. National Science Foundation; 2011. Report No.: NSFDACS06D1412.
3. Horizon 2020—Work Programme 2016–2017–3. Marie Sklodowska-Curie Actions. European Commission; 2016 Mar. Report No.: C(2016)1349.
4. Research Performance and Economic Impact Report 2013/14 [Internet]. EPSRC; 2014. https://www.epsrc.ac.uk/newsevents/pubs/economicimpactreport1314/.
5. Choi BCK, Pak AWP. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. Clin Invest Med. 2006; 29: 351–364. PMID: 17330451
6. Gu J, Bourne PE. Ten simple rules for graduate students. PLoS Comput Biol. 2007; 3: e229. https://doi.org/10.1371/journal.pcbi.0030229 PMID: 18052537

7. Van Noorden R. Interdisciplinary research by the numbers. Nature. 2015; 525: 306–307. https://doi.org/10.1038/525306a PMID: 26381967

8. Marino J, Stefan MI, Blackford S. Ten simple rules for finishing your PhD. PLoS Comput Biol. 2014; 10: e1003954. https://doi.org/10.1371/journal.pcbi.1003954 PMID: 25474445

9. Thessen A, Anne T, David P. Data issues in the life sciences. Zookeys. 2011; 150: 15–51.

10. Knapp B, Bardenet R, Bernabeu MO, Bordas R, Bruna M, Calderhead B, et al. Ten simple rules for a successful cross-disciplinary collaboration. PLoS Comput Biol. 2015; 11: e1004214. https://doi.org/10.1371/journal.pcbi.1004214 PMID: 25928184

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules to make the most out of your undergraduate research career

**Megan Yu[1]*, Yu-Min Kuo[2]**

**1** University of Virginia, Charlottesville, Virginia, United States of America, **2** Department of Cell Biology and Anatomy, College of Medicine, National Cheng Kung University, Tainan, Taiwan

* ymeganyu@gmail.com

Check for updates

OPEN ACCESS

**Citation:** Yu M, Kuo Y-M (2017) Ten simple rules to make the most out of your undergraduate research career. PLoS Comput Biol 13(5): e1005484. https://doi.org/10.1371/journal.pcbi.1005484

**Published:** May 4, 2017



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have read the journal's policy and have the following conflicts: MY is a blogger at the PLOS Early Career Researcher (ECR) Community blog. No other competing interests exist.

In 2008, the Council on Undergraduate Research (CUR), a national organization founded in 1978 that promotes research opportunities for faculty members and undergraduates, featured 2,800 presenters in their annual undergraduate conference. Today, it has developed to include numerous disciplines ranging from biochemistry to theater and drama, and nearly 10,000 members and over 900 universities have participated in its endeavor to promote undergraduate research [1]. These statistics not only highlight the prevalence of undergraduates participating in research but also demonstrate the importance of research in undergraduate education.

Many undergraduates have reported numerous benefits from participating in research. In a study involving about 4,500 undergraduates that participated in undergraduate research opportunities sponsored by the National Science Foundation, respondents reported an increased level of understanding, resilience, and confidence in performing research and motivation to apply for graduate school programs [2]. In another analysis of 76 student interviews from four liberal arts colleges, undergraduates believed they have gained more laboratory (lab) techniques and have developed an attitude to "thinking and working like a scientist" [3]. These lab techniques and research attitudes are essential, as they help undergraduates develop better research habits and the solid foundation of knowledge and experience needed for their future research careers. For instance, knowing how to manage large datasets effectively, such as large patient genetic datasets and electronic health records, and designing proper algorithms and computational models to analyze data are essential skills for undergraduates interested in computational biology. In addition, unlike classroom learning, undergraduate research provides hands-on experience that allows undergraduates to gain a deeper understanding of the scientific process and to develop better research habits.

Despite the multiple benefits that research offers, undergraduates sometimes struggle and feel overwhelmed with the research process. Some undergraduates may not be familiar with the dynamics of the lab and may be afraid to interact with their lab colleagues and mentors. Other undergraduates may not completely understand the purpose of their work and feel overwhelmed by not knowing the results of their experiments before performing them. These consequences could, in turn, have detrimental effects on the relationship between undergraduates and their lab colleagues and decrease the motivation for undergraduates to pursue research in the future [4–5]. In light of these concerns, we propose ten simple rules constructed from our experiences as a college senior and a professor who has worked with undergraduate researchers that would help undergraduates enjoy and intellectually enrich their research experiences. Although this article may have components that are covered elsewhere [6–10], it extends and refines some advice from earlier articles so that they are more suitable for undergraduates.

## Rule 1: Start early

As an undergraduate, you may not know what type of research project you would like to pursue or whether it fits into your future research career. Therefore, it is essential to start early to explore and develop your research interests and goals. Your goal could be to gain more research experience before attending graduate school or to determine whether you prefer working in the industry to working in academia. Or, you might be new to research and hope to determine whether you would incorporate it into your future career or not. Whatever your reason is, be sure to start early to give yourself ample time to reflect on your goals and interests.

Finding the right research lab could take more than emailing several professors or research managers; it might require meeting a member of a lab at a conference or taking a tour of the lab to determine if it is the right fit. You might even consider joining professional research societies or research networks at your university to explore which areas are actively shaping the field and network with other researchers. For instance, the International Society for Computational Biology (ISCB) hosts numerous conferences and forums for computational biologists and students to network and promote their scientific research. It also has a career center for students and researchers to find jobs and be recognized for their talents [11]. Additionally, if you expect to publish during your undergraduate research career, you might want to start early to ask other professors and students within your department or look up the publication patterns of the potential mentor's research group on the lab website.

As you become a new member of a lab, you might need some time to acclimate to the new lab environment and determine your commitment to doing research. You may find that life catches you off guard as you start to juggle between classes, jobs, and extracurricular activities, thus causing you to not find enough time to do research. Starting early, such as during your freshman or sophomore year, would provide you with ample time to explore your research goals and interests and participate in meaningful research activities.

## Rule 2: Know your foundational knowledge and skills

When you begin searching for undergraduate research positions, it is helpful to have already taken the recommended courses related to your research experience. Many professors would evaluate your knowledge and competence in a particular field to predict your success in the lab. Having the background knowledge in the research area of your chosen lab will help you understand the science behind the studies and experiments that are performed and will serve as useful foundational knowledge should you decide to pursue an independent research project in the future. For instance, while a computational biology lab might have a variety of lab members each with a different set of skills, such as a statistician, bioinformatician, or a software developer, it is helpful to have taken courses in computer science, programming, statistics, and biology before joining the lab. If your lab participates in a lot of programming activities, you might also consider brushing up on your coding and programming skills and taking a variety of Massive Open Online Courses (MOOCs) in computer science and programming [12–13]. Another way to gain more foundational knowledge is to read as much as you can about the topics pertaining to your chosen research lab from peer-reviewed journal articles, especially the papers that your chosen lab has published, or from popular science magazines. Table 1 lists some useful online resources for undergraduates to gain additional background preparation for their research experiences.

While it is important to have the foundational knowledge before entering a lab, you should also remember to provide yourself enough time to do research (Rule 1). The process of finding the right time to do research can be complicated and may require you to seek additional help.

Table 1. Useful online resources for undergraduates in background research preparation.

| Type of online resource | Relevant websites and URLs |
|---|---|
| MOOCs | Many companies, such as Udacity (https://www.udacity.com/), Coursera (https://www.coursera.org/), Khan Academy (https://www.khanacademy.org/), and edX (https://www.edx.org/), offer free online courses ranging from machine learning to business management. These courses may be helpful for undergraduates planning to do more interdisciplinary or specialized research. |
| Interactive coding and statistics websites | Codecademy (https://www.codecademy.com/), Code School (https://www.codeschool.com/), and Python for Biologists (http://pythonforbiologists.com/) offer free online training in programming and coding. Simple Interactive Statistical Analysis (SISA; http://www.quantitativeskills.com/sisa/), VassarStats (http://vassarstats.net/), OnlineStatBook (http://onlinestatbook.com/2/index.html), and Stat Trek (http://stattrek.com/) offer free online training in statistical analysis. |
| Scientific news and electronic databases | Many popular science magazines and websites, such as Scientific American (https://www.scientificamerican.com/), Discover (http://discovermagazine.com/), and Phys.org (https://phys.org/), cover the latest research findings in science and technology. Ovid (http://www.ovid.com/site/index.jsp), EMBASE (https://www.embase.com/), Web of Science (http://apps.webofknowledge.com/), and Google Scholar (https://scholar.google.com/) are popular electronic databases that allow you to search for many peer-reviewed articles and books. |
| Web forums and blogs | Many web forums, such as ResearchGate (https://www.researchgate.net/), allow you to ask questions and receive answers from experts when you start doing research. Biostars (https://www.biostars.org/) and SEQanswers (http://seqanswers.com/) are popular web forums in computational biology. Blogs, such as those from PLOS (http://blogs.plos.org/), The BMJ (http://blogs.bmj.com/), and Scientific American (https://blogs.scientificamerican.com/), are also popular online resources for learning. |
| International organizations | Many international organizations, such as the Global Organization for Bioinformatics Learning, Education & Training (GOBLET; http://mygoblet.org/training-portal) and the International Society for Pharmaceutical Engineering (ISPE; http://ispe.org/training), offer online training materials in highly specialized fields, such as engineering and computational biology. |

https://doi.org/10.1371/journal.pcbi.1005484.t001

For instance, you might consider discussing undergraduate research with potential mentors or advisers within your department. You might even visit the career center or take some research methodology or independent study courses at your university to determine if you are prepared. While starting research in your junior or senior year reduces the likelihood of publishing, you might have more foundational knowledge from your classes and have a more individualized approach to achieve your research goals and interests. Whatever it is you choose to do, make sure that you exploit the resources around you and give yourself enough time to decide when is the right time to do research.

## Rule 3: Let passion guide your research interests and goals

Like with many things in life, your interests and passions should help guide you to which research projects and fields you would like to pursue. Being interested in and passionate about the subject matter helps alleviate some of the mental and physical burden you may feel when spending countless hours in the lab. Before accepting an undergraduate research opportunity, you should ask yourself the following questions:

1. Is this research opportunity related to my academic interests?

2. What kind of research experience am I looking for, and what do I hope to gain from the experience?
3. How much time am I willing to commit, and what skills do I have that would contribute to this experience?
4. Do the professors whom I work for have similar academic interests as I do?

While many universities host many conferences and discussion forums in which professors, graduate students, and undergraduate researchers present their work, these events are also an opportunity for aspiring undergraduate researchers to meet with presenters and explore their academic interests. Take advantage of them! Exploring the websites of different research labs and other forms of apprenticeship should not be overlooked, as they are opportunities to gauge your interest in those fields and whether the research lab you are interested in is a good fit for you or not.

Moreover, being enthusiastic about the subject matter helps improve the chemistry you have with your lab director or with your research colleagues (Rules 4 and 6). While your colleagues and mentors are always willing to help you, it would make a better impression and would facilitate more dynamic discussions if you care about the topic. Your mentors and lab colleagues are also more motivated to help you with your project.

## Rule 4: Build positive relationships with your lab colleagues

As you become a new member of a research team, it is critical to be familiar with the dynamics of the lab and build good relationships with your research colleagues. Every lab has its own unique qualities. Some labs, such as basic science labs, may have a large team of senior researchers or graduate students performing experiments to investigate certain phenomena and developing assays on biofluid samples. Other labs, such as social science labs, may have a large team of graduate or undergraduate research assistants enrolling human participants to investigate a certain phenomenon. You might even have a research lab that involves a lot of collaborative research partnerships, sometimes international, with other labs. This is particularly true for labs that are largely interdisciplinary in nature or require highly technical equipment and expertise, such as a computational biology lab or a particle physics lab. There are also some labs that involve a small team of professors analyzing historical data, such as those in the humanities. Regardless of what type of research lab you are in, try to analyze the dynamics among the lab members, as this would help you acclimate to the new environment. You should also learn the expectations of your lab colleagues, as it would help you establish good research habits. Should you decide to have your own lab in the future, understanding lab dynamics and building good relationships with your research colleagues would help you understand your future undergraduate trainees and become a better mentor. Additionally, you should always treat your lab colleagues with respect as this would improve your relationship with them. They could serve as future collaborators, connections, or resources, as they may have more experience in certain research areas than you. Having occasional discussions or chats with them is another way to build better relationships with your lab colleagues.

## Rule 5: Keep an open mind and do not be afraid to ask questions

As an undergraduate, you may not be expected to know how to develop a research question that leads to a significant discovery and is feasible to answer within a limited amount of time. Or, you might be working in a large lab with so many open research questions and projects that you may not have the autonomy to develop your own research project. It is thus important to keep an open mind. Try to learn techniques and obtain new knowledge by having conversations with your senior colleagues. You should also allow your research mentor to guide you and give you advice, such as networking opportunities at professional research societies (Rule 1). Remember that learning how to do scientific research takes some time and effort (Rule 7), and your mentor is there to help you formulate your research project and guide you toward answering that question. Even after you have demonstrated competence in the lab, you should still keep an open mind, as there may be moments where you are inspired with a novel idea that may be relevant to your work. For instance, you might read an interesting news article about a study relevant to your research and wish to incorporate it into your project (Rule 10). Or, you might receive some useful advice from a conversation with a lab colleague and hope to include it into your work (Rule 4).

In addition to keeping an open mind, you should not be afraid to ask your senior colleagues any question regarding your research project or a particular research field. Asking questions is a great way to make an impression and foster open communication with your lab colleagues. It also allows you to learn more about a certain project you might not understand or any networking or presenting opportunity (Rule 9) that may be helpful for your future research career.

## Rule 6: Foster open communication with your research mentor and maintain a work/life balance

Research requires a significant amount of your time and energy and may take a mental and physical toll on your health, particularly if you are doing research during the school year. It is thus important to foster open communication with your research mentor. Remember that your mentor is providing you with the time and resources you need to succeed in the lab, so it is essential that you remain honest about your availability and work. Be sure to let your mentor know about your availability and goals working in the lab during the semester. Research should also be equally balanced with other extracurricular activities that you enjoy, as they would help you maintain a good work/life balance and could be helpful for your future career.

Fostering open communication with your mentor also demonstrates your initiative and progress to your mentor. As your research mentors may be busy with teaching and other scholarship endeavors, it is helpful to set up weekly meetings with them to demonstrate your progress and obtain constructive feedback for your work. You will build a stronger connection with your mentor and your mentor will be more likely to help you by writing you a strong letter of recommendation or helping you coauthor a peer-reviewed paper. If you encounter any moment in which the data you have collected do not meet your expectations, you should still discuss your progress with your mentor at least once per week, because you may fall into a vicious cycle in which you work hard to try to produce positive results in vain. During these instances, your mentor may slightly alter your research project so that you would not fall into that trap and lose motivation in doing research.

## Rule 7: Learn research by doing it

An important part of learning the scientific research process is to actually perform the research. Without setting up the experiment and testing your hypothesis properly, you will never know the truth about your research question [14]. As you perform the experiment, you might find an interesting discovery or gain more experience in doing a particular technique. Doing research also helps you develop better research skills and learn how to deal with setbacks. Regardless of what happens after you perform the experiment, try not to grow too attached to your data and do not put much stress on yourself if your study fails to produce significant results. Instead, you should remain confident and learn from your hardships. Be sure to also have open discussions about your results with other scientists or your lab group. While contributing to a peer-reviewed publication is definitely an impressive accomplishment that many undergraduates aspire to achieve, try not to give yourself too much pressure should your research contributions not turn out the way you expected or do not meet the standards required for a peer-reviewed publication.

Another helpful way to learn the scientific research process is to gain different or more diverse research experiences, especially when your interests change or if things do not turn out the way you have expected. These could be in the form of working on a different project in the same lab or transitioning to another lab to develop a different set of skills. For instance, if your goal is to become an experimental biologist to test for particular types of bioactivity of a drug or biomarker at the cellular or molecular level but your research experience thus far has only focused on data mining in large, biological databases, you might consider moving to a lab that focuses on developing high throughput assays that test these biomarkers and drugs. Whatever your choice is, be sure to let your mentor know of your decision and do the necessary background preparation you need to succeed in the next step of your undergraduate research career (Rule 2). You should also thank your former mentor and lab colleagues, as they have invested some time, effort, and resources in you.

## Rule 8: Be organized

Good organizational skills facilitate effective research and help you maintain a healthy lifestyle. Having an organized lab notebook or a folder with essential background research papers is critical for analyzing data or generating new ideas or proposals for your research project. Most importantly, being organized will help you tremendously when you present your results in a symposium or peer-reviewed publication, as it allows you to complete work in a timely manner. Good organizational skills also help you avoid being overwhelmed and overscheduling yourself with additional activities and other scholarly pursuits.

## Rule 9: Find opportunities to present your work

As an aspiring scientist, you should try to find opportunities to present your work. This could range from having an elevator pitch with a committee member to presenting your work at a conference or in a peer-reviewed paper. These opportunities would not only improve your communication and interpersonal skills and publicize your accomplishments but would also allow you to network with other scientists. Many universities host symposiums during the school semester, and some conferences allow undergraduates to submit an abstract for peer-review. In addition, many summer research programs and postbaccalaureate research programs host poster sessions or other conferences at the end of the research session for students to present their work.

## Rule 10: Keep up with the scientific literature

As an undergraduate researcher, it is helpful to keep up with the scientific literature, as it could provide some inspiration for your research project. For example, your research project might involve the role of a certain gene in the development of a disease and you might come across a scientific paper that cited a publicly available genetic database that could be helpful for your project. You might also have recently encountered an issue with your project for which another research group has just found a potential solution. These news items do not have to be peer-reviewed articles; they could include news from a variety of popular scientific news websites or magazines (Table 1). Keeping up with the scientific literature also helps you gain some additional background knowledge and skills you may need to use in your research project (Rule 2).

## Conclusion

Undergraduate research is an essential part of undergraduate education, as it offers many opportunities, ranging from developing the attitude to work as a researcher to networking and collaborating with other scientists. It is also fun and intellectually rewarding, as it allows you to uncover the truth about a phenomenon or develop better methods to investigate how the world works. These benefits are often not otherwise available in undergraduate education. Therefore, undergraduate research should be included in one's undergraduate career if one is interested in pursuing research in the future. We hope that the advice and tips presented in this article will inspire and encourage other undergraduate researchers to enjoy and make the most out of their undergraduate research careers.

## References

1. Council on Undergraduate Research. Washington, D.C. 2017. http://www.cur.org.
2. Russell SH, Hancock MP, McCullough J. Benefits of Undergraduate Research Experiences. Science 2007; 316(5824): 548–549. https://doi.org/10.1126/science.1140384 PMID: 17463273
3. Seymour E, Hunter A- B, Laursen SL, DeAntoni T. Establishing the Benefits of Research Experiences for Undergraduates in the Sciences: First Findings from a Three-Year Study. Sci Ed 2004; 88(4): 493–534.
4. Petrella JK, Jung AP. Undergraduate Research: Importance, Benefits, and Challenges. Int J Exerc Sci 2008; 1(3): 91–95. PMID: 27182299
5. Sheikh ASF, Sheikh SA, Kaleem A, Waqas A. Factors Contributing to Lack of Interest in Research among Medical Students. Advances in Medical Education and Practice 2013; 4: 237–243. https://doi.org/10.2147/AMEP.S51536 PMID: 24235856
6. Erren TC, Cullen P, Erren M, Bourne PE. Ten Simple Rules for Doing Your Best Research, According to Hamming. PLoS Comput Biol 2007; 3(10): e213.
7. Gu J, Bourne PE. Ten Simple Rules for Graduate Students. PLoS Comput Biol 2007; 3(11): e229. https://doi.org/10.1371/journal.pcbi.0030229 PMID: 18052537
8. Moreno E, Gutiérrez J-M. Ten Simple Rules for Aspiring Scientists in a Low-Income Country. PLoS Comput Biol 2008; 4(5): e1000024.
9. Michaut M. Ten Simple Rules for Getting Involved in Your Scientific Community. PLoS Comput Biol 2011; 7(10): e1002232. https://doi.org/10.1371/journal.pcbi.1002232 PMID: 22046114
10. Bourne PE, Barbour V. Ten Simple Rules for Building and Maintaining a Scientific Reputation. PLoS Comput Biol 2011; 7(6): e1002108. https://doi.org/10.1371/journal.pcbi.1002108 PMID: 21738465
11. International Society of Computational Biology. Bethesda, MD. 2017. https://www.iscb.org/.
12. Manallack DT, Yuriev E. Ten Simple Rules for Developing a MOOC. PLoS Comput Biol 2016; 12(10): e1005061. https://doi.org/10.1371/journal.pcbi.1005061 PMID: 27764089
13. Loman N, Watson M. So You Want to Be A Computational Biologist? Nature Biotechnology 2013; 31 (11): 996–998. https://doi.org/10.1038/nbt.2740 PMID: 24213777
14. Schwartz M. The Importance of Stupidity in Scientific Research. Journal of Cell Science 2008; 121(11): 1771. https://doi.org/10.1242/jcs.033340 PMID: 18492790

EDITORIAL

# Ten Simple Rules for Developing a MOOC

**David T. Manallack***, **Elizabeth Yuriev**

Faculty of Pharmacy and Pharmaceutical Sciences, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia

* David.Manallack@monash.edu

If you have not heard of the word MOOC, it refers to **M**assive **O**pen **O**nline **C**ourses, and their appeal has raised the interest of most tertiary institutions. MOOCs have rapidly become "the new black" for online learning [1]. The first course emerged in 2008 from the University of Manitoba (http://cck11.mooc.ca/), and the term MOOC was coined at that time by Dave Cornier (http://davecormier.com/edblog/2008/10/02/the-cck08-mooc-connectivism-course-14-way/). An ever-increasing number of courses from major universities with a range of course providers is now established, and new ones are coming out on a weekly basis.

At Monash University, we recently completed and delivered a MOOC (the "Science of Medicines") through the FutureLearn platform (https://www.futurelearn.com/courses/the-science-of-medicines). The course has been very successful (total enrollment of 28,500 learners) and has run five times. To facilitate the development of MOOCs, we have put together a set of ten simple rules based on our experiences to give you tips on what to look out for and what to avoid. It must be said, however, that this is not an exercise for the faint-hearted, as MOOC development entails a considerable amount of work. Within the world of MOOCs, there are both introductory and intermediate level courses. Our rules primarily relate to the former and are aimed at academics; however, many of the principles are in common.

## Rule 1: Educator Mission—Establish the Purpose

The question that needs to be asked is, "Why do you want to develop a MOOC?" Is it merely fashionable to have one or is there a serious educational imperative? Certainly a major appeal of MOOCs is that they provide education to a multitude of people across the planet (usually without cost). Instead of lecturing to a class of 100 individuals, you can reach many thousands of people, often on a topic close to your heart. One should also keep in mind the broad diversity of participants who will differ considerably in their educational background, age, and culture. In some cases, the MOOC is oriented to showcasing the educational strengths of an institution and represents, in effect, a branding exercise. Academics called in to create the MOOC need to debate the learning objectives and its purpose for their own clarity and motivation.

## Rule 2: What Is a MOOC?—Experience a MOOC Firsthand

How do they operate? One simple way of exploring these online offerings is to enroll in a MOOC and do some reconnaissance. Given that all MOOCs suffer from a rapid drop-off in numbers throughout the course, you need to consider how to maintain the interest of the participants. Involving yourself in a MOOC will let you understand what works and which platforms are appealing.

OPEN ACCESS

**Citation:** Manallack DT, Yuriev E (2016) Ten Simple Rules for Developing a MOOC. PLoS Comput Biol 12(10): e1005061. doi:10.1371/journal.pcbi.1005061

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** October 20, 2016



**Funding:** No funding supported this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Rule 3: Select a MOOC Provider

Going it alone will probably involve a lot of work and likely won't play to the strengths of your organization. As such, you should pin yourself to an existing platform/provider (e.g., Coursera, edX, FutureLearn. For the full list, see https://www.mooc-list.com/). Start by exploring, and focus on what the provider offers to the MOOC developers. For example, do they give advice or assistance on structuring a MOOC, quality control, captioning, and hosting services? Once a provider is chosen, a dialogue can be started to gather details and determine whether there is an alignment between their organization and your educational vision. If legal agreements are needed (e.g., royalty sharing), then an appropriate amount of time should be allocated for procuring advice and finalizing contracts.

## Rule 4: Decide on Subject Matter

If the MOOC is showcasing your institution, then this already narrows the subject area. The MOOC can't teach an entire degree in a matter of a few weeks, so keeping it enjoyable, punchy, and interesting is vital. The decision on the topic needs considerable thought to keep future participants engaged. Again, the educational vision needs to drive this to carefully select the material to deliver. Personal passions to teach the world everything close to your heart will need to be curtailed and put in terms of "what specifically would enlighten them." The mission, objectives, and specific structure of the MOOC should be carefully defined. This will also help educators who will be invited to design the MOOC to buy into the development process.

## Rule 5: Determine Governance

From the earliest murmurings of your MOOC development, several people will have been involved. For efficiency, there needs to be an established team with clearly defined roles. This will include several academics to both write and present material, a videographer and video editor, a graphics artist, project manager, text editor, and solicitor (where needed). Budgets and schedules must be assembled including tasks to be completed, meeting dates, etc. Even more important is communication, via email and shared document areas. MOOC providers can be extremely useful to map out tasks, ensure quality control, monitor intellectual property, facilitate dialogue, and set up deadlines. The lines of communication also need to be stated, such as who will liaise with the MOOC provider. Having a release date certainly sharpens the mind.

## Rule 6: Design Your MOOC

Clearly the design task itself is a huge topic and cannot be fully covered in this set of rules. Depending on the MOOC provider, they are likely to have a set format for their courses. This is an advantage, as it specifies valuable guidelines on videos, quizzes, discussion boards, polls, and so forth. Each person generating the teaching material then works to a common framework. This not only benefits the educators producing the material but greatly helps learners who get used to a particular style. Estimated time commitment for the learner is another vital consideration for the design to render the MOOC enjoyable and achievable.

One more thing to consider is what will set your MOOC apart from the rest. What is special about your team? For example, in our "Science of Medicines" MOOC, we have used videos of characters inhabiting a fictional village named Pharmville to start a conversation about a range of ailments and corresponding treatments. Learners could relate to these characters and were motivated to participate in discussions sharing personal experiences. We also decided not to solely use a standard multiple-choice question approach to get learners to revise. Instead, we invited them to do crosswords, and this turned out to be an engaging and effective revision technique.

Finally, the design should consider the level at which the course is pitched. MOOC participants will have a wide educational background, and with such a diverse set of people, there is a need to carefully explain concepts, provide additional resource material, and avoid jargon. If not—they will soon move on. The language used and delivery style must be well crafted and edited by a professional to ensure it is consistent and understandable. Captioning is another important resource for non-English speaking participants or those with hearing impairments and will require careful checking to avoid errors. Employing professional videographers with graphics skills also facilitates the entire process. If you are filming multiple sections of the MOOC on one day, it is advisable that you have multiple changes of wardrobe so that keen-eyed participants notice the change of clothes.

## Rule 7: Pilot Test Your MOOC

Quality control is the job of all people involved and will be heavily scrutinized by the provider. Consistency of style, graphics, and cross-examination of material will help, but errors are always made. The aim of pilot testing is to pick up problems before releasing the MOOC to the public. By taking a global view of a MOOC, the pilot test (preferably by lay people) can get an overall idea of the course. Not only does this pick up inconsistencies and errors, it can also gauge the overall merit of the course.

## Rule 8: Promote the MOOC

The MOOC provider will be your partner here to promote the MOOC with an international outlook. Within your own institution, there must also be a strategy to inform people at a national level with press releases and through alumni networks. Given the volume of work that went into developing the MOOC and the passion involved, you need bang for your buck. Promotion also harks back to the purpose of the course. Is it to showcase the institution or is it about education for the masses? Or both? Either way, the word needs to get out there about this fantastic new MOOC.

## Rule 9: Manage the MOOC

The moment you launch your MOOC, participants will start "talking." The discussion forums will need to be moderated, and other activities will require attention. Communication channels and contact people need to be fully functional at this time. Some elements of the MOOC may require ongoing refinement. But more importantly, questions from learners and debates between them will likely arise. This will require further moderation and regular feedback. Educators, responsive and eager to provide feedback, make for happy and appreciative learners to avoid the MOOC appearing "dead." And remember, there is always someone on the planet that is wide awake and learning from your MOOC—it is a 24 hr/day enterprise.

Another foible is that the developers, as experts in their field, may have inadvertently introduced difficult concepts that require further clarification. By quickly putting together short videos while the MOOC is running, these misunderstandings can be mitigated (e.g., the YouTube channel associated with our MOOC https://www.youtube.com/channel/UCGp8ENReT_kw6YRPm33MIVQ).

## Rule 10: MOOC Postmortem—Debrief

As the MOOC approaches the release date, the MOOC provider can continually communicate enrollment numbers. They will ultimately inform the institution about various metrics such as demographics, length of time in the course, completion rates, and so forth. These can be used to compare against other offerings and be discussed internally by your MOOC team. Data concerning the demographics of the participants at the start and the end of the course provides an insight into those who persist through the entire MOOC. Of interest is their prior educational experience, which can help in the fine tuning of subject areas that were found difficult to assimilate.

Of prime importance is the need to recognize the involvement of educators, as these exercises take up a considerable amount of time that detracts from other activities. Lastly, the MOOC doesn't die there. If successful and well received, it is expected that the MOOC will be rerun periodically, which again means moderators need to be summoned.

Incorporating MOOC materials into your regular courses is another benefit following its development. Moreover, the resources can be used to develop graphics material and video vignettes to promote your courses to future undergraduates.

Overall, developing a MOOC is onerous; but in a team environment and with a structured framework, they provide excellent motivation for creating engaging teaching materials designed to enlighten as many as possible.

## Acknowledgments

The authors would like to acknowledge the editors, writers, presenters, and videographers from the "Science of Medicines" MOOC team: Dan T. Malone, Kirstie J. Galbraith, Ian C. Larson, Keith Sewell, and Jenny M. McDowell, Faculty of Pharmacy and Pharmaceutical Sciences, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville VIC 3052, Australia.

## Reference

1. Pappano L. The Year of the MOOC. The New York Times. November 4, 2012, ED26. Available at: http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html.

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Finishing Your PhD

**Jacopo Marino[1¤*], Melanie I. Stefan[2], Sarah Blackford[3]**

1 Department of Chemistry, University of Zurich, Zurich, Switzerland, 2 Department of Neurobiology, Harvard Medical School, Boston, Massachusetts, United States of America, 3 Society for Experimental Biology (SEB), Lancaster University, Lancaster, United Kingdom

## Introduction

After years of research and with completion in sight, the final year of the PhD often represents the most challenging time of a student's career, in which the ultimate reward is the PhD honor itself. A large investment in time, energy, and motivation is needed, with many tasks to be completed; concluding experiments must be carried out, results interpreted, and a research story mapped out in preparation for writing the final thesis. All the while, administrative obligations need attention (e.g., university credits and mandatory documents), papers may need to be published, students mentored, and due consideration paid to planning for the next career move. Without some form of strategic action plan and the employment of project management skills, students run the risk of becoming overwhelmed and run down or of not meeting their final deadlines. Personal time management and stress resilience are competences that can be developed and honed during this final period of the PhD.

Here, we present ten simple rules on how to deal with time issues and conflict situations when facing the last year of a PhD in science. The rules focus on defining research goals in advance and designing a plan of action. Moreover, we discuss the importance of managing relationships with supervisors and colleagues, as well as early career planning.

## Rule 1: Plan Your Last Year in Advance

Preparing a plan of action for the final year of your PhD is vital. Ideally, devised and agreed upon with your supervisor, a plan will help to optimize the time left and reduce feelings of being overwhelmed. Individuals plan in different ways; some prefer to work towards their goals in a stepwise linear fashion, whilst others are more comfortable flitting from task to task until all the jobs are done. There is no definitive way to plan, so find out what works best for you. You may decide to map out a timeline, or perhaps a mind-map is your preferred planning style. Whichever method you use, it's important that you adhere to your plan whilst allowing for some flexibility (but not distraction or procrastination).

Your time frame will vary according to the organization of your graduate school, your supervisor or advisory committee, and even your graduation date, but one year before submission of your doctoral thesis is the time when you should decide on how best to invest the last months of your research and associated activities. Having a plan of action will help to avoid time wasting, e.g., being distracted by superfluous experiments that might be interesting but are not necessary. Furthermore, from a psychological point of view, referring to a concrete plan can make you feel more secure and in control. Ideally, the supervisor and PhD student should both agree on the overall plan (with provision for the unexpected, e.g., technical issues), with intermittent reviews every few weeks to check that progress is being made. Your supervisor should also be able to advise you on the organization and writing of your thesis—for example, its structure—and the number and length of chapters to include.

## Rule 2: Make Your Priorities Clear

Select the activities you want to include in your plan. What are your priorities? They are likely to include experiments that will give the thesis a conclusion or that may be necessary to publish a final paper. Mandatory administrative tasks will also need attention, and allowing time to prepare for your next career move will give you the best chance of a seamless and successful transition post-PhD. As a final year PhD candidate, you are likely to have acquired high-level competencies comparable to those of a junior postdoctoral researcher, in which case your supervisor may offer you responsibility for new projects or graduate students. Saying no to him/her can be difficult for various reasons, e.g., fear of potentially creating conflict in your relationship or causing a negative reaction or of perhaps losing the opportunity to be included in future research activities and publications. It can also be difficult to let go of a topic or project to which you are wedded or to miss out on the opportunity to help train the next generation of scientists. In such situations, referring back to your plan (Rule 1), previously agreed upon with your supervisor, should help to remind you both of your priorities and deadlines, making negotiation easier. However, should any conflict of opinion arise between you, bear in mind that finding a mutually agreeable solution is the best way forward. You can take advice from a mentor or refer to the many publications that provide approaches and tactics for effective negotiation. If the relationship between you and your supervisor is more complicated and cannot be resolved by a discussion, you may need to turn to your graduate school, your academic committee, or other senior managers in your institution, who can act to mediate the situation.

## Rule 3: "The Truth Can Wait"

A research project is never really finished, so do not try to do everything

**Citation:** Marino J, Stefan MI, Blackford S (2014) Ten Simple Rules for Finishing Your PhD. PLoS Comput Biol 10(12): e1003954. doi:10.1371/journal.pcbi.1003954

**Editor:** Philip E. Bourne, National Institutes of Health, United States of America

**Published** December 4, 2014



**Funding:** The authors have received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* Email: jacopo.marino@chem.uzh.ch

¤ **Current address:** Gene Center, Department of Chemistry and Biochemistry, University of Munich, Munich, Germany

before submitting. In fact, the perfect doctoral thesis does not exist; there are students with good research projects and many publications and others with more difficult and testing challenges who are still waiting for their first paper. If the project is ambitious, it might take several years to reach the final goal, and thus the thesis might only be a small part of the whole story. If the project is going well, it will open up new research questions and future directions, some of which will be beyond the scope of a PhD. At some point, you need to decide that what you have is enough for a PhD and start writing (a strategy we heard described at a dissertation-writing seminar in Cambridge as "the truth can wait"; it helps to write this on a post-it note and stick it on your computer!). Starting to write the thesis is not easy when there is a sense that more could be done to accumulate more data and a fuller story; a common mistake is to go back to the lab instead of getting started with the results chapters of the thesis. To postpone writing will cause delays and not necessarily improve the thesis whilst increasing the prospect of unfulfilled and extended deadlines. Thus, once the experiments that you have agreed on have been completed, it is really important to start writing with the data in hand.

## Rule 4: Enlist Support

Finalizing experiments and writing the thesis (and even papers), as well as considering your next career transition, can be stressful and even isolating. It is a contrast to the relatively more relaxed earlier years of the PhD experience, and the writing process does not come naturally to everyone. The prospect of facing these stresses alone can make the experience even harder to bear, so it is advisable to communicate with and find support in those you trust and respect. Relying on such people during this period can help to ease the strain and enable you to achieve your final aims so that you arrive at your PhD graduation with your sanity still intact! Talking about personal feelings with selected colleagues usually helps you to realize that you are not alone, whatever difficulties and challenges you might be experiencing with your research project, supervisor, or coworkers. Sharing uncertainties and talking through issues can be constructive, helping you to understand the strategies other people use to cope with similar problems. As well as colleagues, it can also help to talk to friends and family, even though they won't be as au fait with the highly particular challenges you are experiencing. You can share your feelings and anxieties with them, but they can also act as a welcome distraction to help you to relax and take a break from thinking about the stresses of your PhD.

Support and advice can also come in the shape of courses, books, blogs, mentoring, etc. There is much published on the subject of how to write a thesis [1]. Furthermore, graduate schools, such as those in which we are based, usually offer courses to help PhD candidates improve their personal and professional skills. For example, the University of Zurich organizes courses on, amongst others, time and self-management skills, managing conflict, and academic writing and publishing [2]. The Graduate School of Arts and Sciences at Harvard lists workshops and resources offered across the university on topics such as scientific writing, time management, and overcoming procrastination. In addition to relying on your supervisor, postdoctoral researchers in your group or department (or even friendly collaborators) may agree to read chapters of your thesis and comment on aspects such as content, logical flow of ideas, and the overall structure. At a later stage, you may want to engage someone to check your grammar, spelling, and reference style (this can be especially important if you are not writing in your native language). If your PhD defense includes a presentation, try to practice beforehand, preferably in front of some of your peers, and include asking for feedback and possible questions that may come up. This should make you feel more prepared and confident.

## Rule 5: Get Familiar with the Software

Being familiar with software for both writing and making figures will facilitate the creation of your thesis. One of the most effective tools with which to produce a scientific document is LaTeX (www.latex-project.org). This software, freely available, is not as immediately understandable as other text editors, but the advantages are greater: it offers a professional layout similar to published books, it makes the insertion and management of figures easier as their position in the file does not depend on text editing, and it allows for easy typesetting of mathematical equations and referencing of articles from a bibliography database. Moreover, the text file size does not increase while inserting figures, making its handling easier. An example LaTeX package for typesetting dissertations is "classicthesis", written by André Miede (http://www.miede.de/index.php?page=classicthesis). Although advantageous, LaTeX can also present disadvantages. In contrast to commonly used text editors (e.g., Microsoft Word), it does not make it easy to track changes in the manuscript, often a preferred way for supervisors to correct theses in an electronic form. Therefore, we suggest you discuss the preferred software with your supervisor when you agree upon your plan (Rule 1).

A professional design software can also speed up the creation of figures for your thesis, which can be further used for your final PhD presentation, so check whether your institution provides an introductory course to some of these software packages. Taking a one-day class can save you a lot of time later. Organize your bibliography; many excellent reference managers exist that allow you to catalogue and annotate the papers you have read and integrate them seamlessly with text processing software (e.g., Endnote or the freely available Mendeley and Readcube). Choose one that fits your needs and check whether your university provides institutional licenses (and be disciplined about adding each paper you read to it!).

Consider using version control software. This allows you to keep a log of all the changes you make to a file or directory and makes it easy to recover a previous version if something goes wrong or to merge two versions of a file. This is often used in software projects to produce, document, and improve computer code, but it can also be useful when working on a long text document, such as a dissertation. Commonly used free version control systems include git/GitHub [git, github], Subversion [svn], and Bazaar [bzr] (see Table 1).

Most important of all is to have a backup strategy. A hard-drive crash at the wrong moment can set your work back by weeks and jeopardize the timely completion of your thesis. Institutions or departments will often have a backup system employees can make use of. This may require you to install a specific piece of software on your computer that backs up your data at regular intervals or to save your file on an institute server. Contact the information technology (IT) department at your institute to learn about your options.

## Rule 6: Know Your University's Procedures and Regulations

During the course of your PhD, you will have been acquiring project management skills, such as organizing your time and resources, reviewing progress, and meeting deadlines. In order to avoid last-minute

**Table 1.** List of free version control systems.

| Version Control System | Developer | Available: |
|---|---|---|
| [jabref] | JabRef Development Team (2014) JabRef [Software] | http://jabref.sf.net |
| [git] | Git Development Team (2014) Git [Software] | http://git-scm.com/ |
| [github] | GitHub Development Team (2014) GitHub [Software] | https://github.com/ |
| [svn] | Apache Software Foundation (2014) Subversion [Software] | https://subversion.apache.org/ |
| [bzr] | GNU Project (2014) Bazaar [Software] | http://bazaar.canonical.com/en/ |

doi:10.1371/journal.pcbi.1003954.t001

surprises, you can capitalize on and develop these skills during the final year of your PhD. Prepare a list of all the documents and certificates that you will need, even before you start writing; it will be of critical importance to include this information in your plan and priorities (Rules 1 and 2). Having a good working relationship with someone who can help you to navigate a bureaucratic process will usually be an asset and will ensure you are familiar and aware of all the rules. Considering the amount of documents and certificates that are needed for handing in a thesis, it is advantageous to introduce yourself to the institute secretary or human resources manager, as well as any other staff who can help you to deal with the administrative side of the process. Don't rely on previous documents, which may have been revised since the last person in your group graduated. Be aware of all the necessary institutional administrative requirements (e.g., credit points, research seminar attendance, publications, etc.), as well as the faculty criteria, including deadlines (as well the date of the graduation ceremony), thesis copy numbers and format, font size, binding, and supporting documents. Take time to go through the list of documents and start collecting them in a folder. Get the formatting right early on, e.g., by using a dedicated template file. With your documents in order, you are bound to feel you have the situation more under control, which can help to reduce stress and enable you to focus more closely on writing your thesis.

## Rule 7: Exploit Synergies

You are doing a lot of work for your thesis, so use it to your advantage. The literature review in your introduction can also be used to write and publish a future review article, an idea that might also be welcomed by your supervisor. If you are intending to write a grant proposal for a postdoctoral fellowship on a similar research topic, you can use some of the thesis introduction and future directions as a basis for your research plan. If you are keen to gain teaching experience, you could propose a short course on your specialty area. For instance, at Harvard Medical School, senior graduate students and postdoctoral researchers can be involved in lecturing on short, specialized "nanocourses" [3]. You may also be able to deliver a specialized lecture within a class your supervisor is teaching or, ideally after you have completed the PhD, teach at a workshop or summer school.

Take advantage of opportunities to deliver a talk as an invited speaker at a conference or at another institute, for example, if you are visiting a research group or investigating possible postdoctoral options. This will give you the chance to practice your defense presentation in front of an unfamiliar audience and, at the same time, allow a potential future supervisor and colleagues to gain a more complete picture of your research interests, skills, and personality.

## Rule 8: Pay Attention to Your Career

It is not always easy to decide on which career path to follow after your PhD. You have been trained primarily towards an academic research career, and so many PhD graduates choose to continue on with a postdoctoral position as their first career destination. This is perfectly acceptable, and many industrial employers look upon early-career postdoctorals favorably. However, it is worth bearing in mind that permanent tenured positions are hard to secure nowadays and competition is tough, with less than 5% of those who complete a PhD ultimately realizing an academic career [4]. For those who are determined to have an academic career, a strategic research plan is crucial; for those who are unsure, a viable alternative career plan is equally important.

Knowledge of your professional and personal skills and capabilities, personality, values, and interests, as well as how to map them onto the job market and sell them to employers, will help you to make effective career decisions and a successful transition to your next job. In addition, factors such as your personal situation and priorities, mobility, and preferred work–life balance all need to be taken into consideration before entering the complicated world of the job market. Be ready to make compromises either in your work or personal life, depending on your priorities. Take advantage of courses and professional career guidance and coaching while you are still at university, as they are usually offered free of charge. Along with books and websites, face-to-face career support can help raise your self-awareness and knowledge of the job market so you can start to decide which types of career may best suit you. Blackford's book and blog [4] contain useful material on career planning for bioscientists, with concrete examples of different career paths within and outside of academia, and further information and resources. In addition, the *Science*Careers portal offers an online tool [5] to create an individual development plan and explore your career options based on your skills, interests, and values. Also, take advantage of dedicated career job boards associated with specialist websites, such as that of the International Society for Computational Biology [6].

How soon should you start job seeking? Finding a job whilst writing up your thesis can seem like an attractive prospect, but it's important to consider that applying for jobs can easily take up as much time as working a full-time job. Then, if you do secure a job, the time left for writing up your thesis, completing experiments, and wrapping up your lab work will be seriously limited. It is exceedingly hard to write a doctoral thesis in the evenings after work or on the weekends, so in case you are offered a job before you have finished the PhD, consider seriously how this might affect your work and life. On the other hand, finishing a PhD when scholarship money has been seriously reduced (or has run out) comes with a different set of challenges. Many students need to tap into their savings (if indeed they have any), drastically reduce their spending, and move out of their accommodation. Losing employment at the university can also affect health insurance, social security, and visa status. Finishing up a PhD under these additional constraints and pressures can be extremely challenging, both logistically and psychologically. To ensure that you can concentrate all your time on (and get paid for) finishing your PhD, start planning ahead one year earlier. Be aware of your university's regulations, talk to your supervisor about the funding situation (is it possible for you stay on as a postdoctoral researcher for a short period?), and know what you need to do in order to finish on time (Rule 1).

## Rule 9: Network

Unofficial statistics tell us that only around 30% of jobs are advertised, so to enhance your employment prospects you would be well advised to network in order to access the hidden job market. During the final year of your PhD, and even earlier, you can build up and extend your network so that your chances of finding the job of your choice are optimized. If you are looking for research positions, your supervisor might have contacts or know about positions available in academia or industry. Reviewing your personal network further will reveal it consists of colleagues, friends, and family. You may also have a wider network of collaborators (research and industry), people associated with your research whom you have met during the course of your PhD, as well as many others. Conferences, seminars, informal gatherings, and learned societies are great places to meet the academic community face to face or to broaden your horizons. Job fairs are held at universities and sometimes during conferences, where experts from industry look for potential employees as well as sometimes provide informal advice on your curriculum vitae (CV). Try to exploit these opportunities if they come your way.

A relatively recent, and highly democratic, addition to the networking system is social media, through which it is possible to meet people online from all over the world and from all walks of life. More and more professors, researchers, students, policy makers, science "celebrities", science communicators, industry personnel, and professionals have a presence on social media, using it primarily for work-related purposes. Researchgate, LinkedIn, and Twitter are probably the most useful platforms for networking with academia, business, and the wider world, respectively. Your online profile should be fully completed and reflect your expertise, achievements, and personality. Used to greatest effect, social media will give you access to information, jobs, and influential people—its importance to you as a PhD student cannot be overestimated.

## Rule 10: Leave on Good Terms

Wrap up the work in your lab, especially if you are leaving the institute. This includes any required training of new personnel in the methods and techniques you use, having lab notes in order, making it easy for other lab members to access your protocols and data, organizing and labelling your reagents and equipment, and documenting your computer code. If someone is taking over an unfinished project from you, take time to hand it over. Discuss with your supervisor to find a solution for who will do the final experiments, how to proceed with the writing of journal manuscripts, and what should be the order of authorship. If you have started a project that you want to take with you to your new lab, discuss with your supervisor how to handle possible future publications and how to agree on material transfer. If your work resulted in patents or patentable innovations, make sure you are clear about regulations concerning patents and intellectual property, both at your PhD institution and at the institution to which you are moving. Stay in touch with your former colleagues and cultivate the contacts you have made in graduate school; they are sure to be useful during the course of your career.


## Acknowledgments

Jacopo Marino is grateful to colleagues from the University of Zurich for the everyday discussions that have inspired this manuscript. Melanie I. Stefan is likewise grateful for discussions on the topic with fellow predocs (and sympathetic postdocs) at the European Bioinformatics Institute. She would also like to acknowledge advice and support from Nicolas Le Novère and Susan Jones, which helped her navigate her PhD and graduate in a timely manner. She has since learnt a lot from discussions with colleagues at the California Institute of Technology, the University of Tokyo, and Harvard Medical School.


## References


1. Turabian KL (2013) A Manual for Writers of Research Papers, Theses, and Dissertations. 8th edition. Chicago (Illinois): The University of Chicago Press.
2. University of Zürich (2014) Courses for PhD candidates and postdocs. Available: http://www.grc.uzh.ch/phd-postdoc/courses-uzh_en.html. Accessed 30 October 2014.
3. Bentley AM, Artavanis-Tsakonas S, Stanford JS (2008) Nanocourses: A Short Course Format as an Educational Tool in a Biological Sciences Graduate Curriculum. Cell Biol Educ 7: 175–183.
4. Blackford S (2013) Career planning for research bioscientists. Wiley-Blackwell. Available: http://www.biosciencecareers.org. Accessed 30 October 2014.
5. Hobin JA, Fuhrmann CN, Lindstaedt B, Clifford PS (2012) You Need a Game Plan. Science Careers Career Magazine. Available: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2012_09_07/caredit.a1200100. Accessed 30 October 2014.
6. International Society for Computational Biology (2011) ISCB Careers. Available: https://www.iscb.org/iscb-careers. Accessed 30 October 2014.

Editorial

# Ten Simple Rules for Developing a Short Bioinformatics Training Course

**Allegra Via[1], Javier De Las Rivas[2], Teresa K. Attwood[3], David Landsman[4], Michelle D. Brazas[5], Jack A. M. Leunissen[6], Anna Tramontano[1], Maria Victoria Schneider[7]***

**1** Biocomputing Group, Department of Physics, Sapienza University of Rome, Rome, Italy, **2** Bioinformatics & Functional Genomics Research Group, Cancer Research Center (IBMCC, CSIC/USAL), Salamanca, Spain, **3** Faculty of Life Sciences and School of Computer Science, University of Manchester, Manchester, United Kingdom, **4** Computational Biology Branch, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Maryland, United States of America, **5** Informatics and Bio-computing, Ontario Institute for Cancer Research, MaRS Centre, Toronto, Canada, **6** Laboratory of Bioinformatics, Wageningen University, Wageningen, The Netherlands, **7** Outreach and Training Team, EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom

## Introduction

This paper considers what makes a short course in bioinformatics successful. In today's research environment, exposure to bioinformatics training is something that anyone embarking on life sciences research is likely to need at some point. Furthermore, as research technologies evolve, this need will continue to grow. In fact, as a consequence of the introduction of high-throughput technologies, there has already been an increase in demand for training relating to the use of computational resources and tools designed for high-throughput data storage, retrieval, and analysis. Biologists and computational scientists alike are seeking postgraduate learning opportunities in various bioinformatics topics that meet the needs and time restrictions of their schedules. Short, intensive bioinformatics courses (typically from a couple of days to a week in length, and covering a variety of topics) are available throughout the world, and more continue to be developed to meet the growing training needs. The challenges, however, when planning, organising, and delivering such courses, are not trivial [1], especially considering the heterogeneous backgrounds of participants. Here, we address such challenges and present a consensus of rules derived from the shared expertise of several bioinformatics trainers. While the rules apply broadly to bioinformatics training, aspects addressing specific audiences are also discussed in order to make these rules pragmatic and applicable to a wide range of readers. Delivering bioinformatics training is both crucial to facilitate the use of, and to exploit the investment in, bioinformatics tools and resources, and an excellent opportunity to solicit user evaluation and feedback to improve them. One point of crucial interest to the training course community concerns material preparation and distribution. Preparing effective materials (slides, notes, references, etc.) entails a huge effort that would be enormously facilitated if course developers could start from a body of available materials, for example if they could gain access to repositories of materials deposited by trainers of other courses. This was one of the reasons motivating the Bioinformatics Training Network (BTN) to set up the BTN website (http://www.biotnet.org/), which has been planned as a vessel for the training community to share and disseminate course information and materials. Course developers are warmly welcome to subscribe to the site and make available their materials to the community [2].

## Rule 1: Set Practical and Realistic Expectations

It is critical to explicitly identify the training objectives and expected outcomes from the outset. Begin by devising the title of your course and specifying the target audience (e.g., laboratory biologists, computational scientists). This information is not only useful for trainers to help appropriately focus and weight the contents of their training sessions, but is also vital for participants. By explicitly stating the course objectives up front, trainees are better oriented to the expected outcomes and are more likely to be satisfied with the course. As most training sessions are based on slide presentations, dedicate at least one slide (preferably, while providing the session overview) to the learning objectives, and mention how these will be achieved, using specific examples whenever possible; if appropriate, also mention how the knowledge gained and skill set(s) will be useful for trainees' work environments. Stating what participants will not learn to do (e.g., to avoid over-estimation of the depth of analysis that can be achieved in a short course) is also important for tempering their expectations.

## Rule 2: Verify That Trainees' Expectations Match Course Scope

Verify that trainees' expectations match what will be delivered. The most effective mechanism to ensure that expectations are well matched is to collect information from trainees prior to the training session itself (e.g., via a questionnaire), or by discussions with trainees at the start of the course. Obtaining such information early on allows time to alter course materials to better meet participant expectations, for example by adjusting case studies and

**Citation:** Via A, De Las Rivas J, Attwood TK, Landsman D, Brazas MD, et al. (2011) Ten Simple Rules for Developing a Short Bioinformatics Training Course. PLoS Comput Biol 7(10): e1002245. doi:10.1371/journal.pcbi.1002245

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** October 27, 2011



**Funding:** This work was partly supported by the Intramural Research Program of the NIH, NLM, NCBI, and by funds awarded to the EMBL-European Bioinformatics Institute by the European Commission under SLING, grant agreement number 226073 (Integrating Activity) within Research Infrastructures of the FP7 Capacities Specific Programme EMBL-EBI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: vicky@ebi.ac.uk

examples to reflect the audience's interests. Furthermore, this will make you aware of the trainees' different backgrounds. Read, or listen to, and evaluate all responses, both to discern whether the course content matches participant expectations and to learn what the trainees' needs are. Such information will also allow you to detect clusters of trainees: e.g., those working with a particular model organism, those more interested in DNA than in proteins, or more plant than animal scientists. Useful information to collect includes their research backgrounds and computational skill sets, their current projects relevant to the course, and their expectations of the training (e.g., what reasons led them to apply for this particular course?). Also solicit information from trainees about the biological problems they wish to solve by participating in the course.

## Rule 3: Plan Exercises and Activities and Test Resources before Delivery

Plan the course in independent units/modules, each with an introduction, set of aims, list of actions, and potential difficulties. When a new module is introduced, recall the achievements of the previous module, and state what tasks participants will be able to additionally accomplish at the end of the new module.

If you, the trainer, are also responsible for the resource/tool being presented, you are likely to be able to handle unexpected queries or problems. However, many trainers deliver sessions on resources/tools built and maintained somewhere else by someone else, using someone else's data. Regardless, always prepare an alternative plan in anticipation of unforeseen difficulties. For example, at short notice, you might not be able to use live queries, so ensure that you have sufficient back-up material (e.g., animations, videos, etc.) to allow you nevertheless to deliver your training session effectively.

To appear as prepared and experienced as possible, try your practical exercises beforehand. In cases where the query or task required to a bioinformatics server takes a long time, or is too demanding on the service provider, either begin with smaller query datasets, or provide the task results after trainees have prepared the query set-up, so that they still gain the experience of performing the task and class time is used more efficiently. It is important to note that some service providers will often hold query results for 48 hours.

## Rule 4: Ensure Computational Equipment Preparedness and Hands-On Support Availability

Ensure (or rather, insist) that workstations (Linux, Mac, or PC) have all the necessary software installed to allow trainees to complete the course. Make sure that the venue provides each trainee (or, at most, each pair of trainees) with one computer. Where trainees are required to bring their own workstation (e.g., laptop), provide enough instruction and test commands to ensure that software and dependencies have been properly set up ahead of time. Request that a system support technologist be available, and in the room, when starting your sessions, to ensure the functionality of the classroom workstations and/or of the participants' personal computers.

Do not underestimate the trainer/trainee ratio, especially in consideration of the trainees' diverse backgrounds. Be prepared to provide extra hands-on support while trainees become familiar with new interfaces, tools, and resources. Such support may be provided by trainers of other modules, tutorial assistants, past trainees, or even current trainees who are familiar with the tool/resource basics.

## Rule 5: Use the Dynamic World of Bioinformatics Resources and Tools as a Learning Opportunity

Provide time references for the information you deliver, as bioinformatics resources and tools, and stored data, evolve continuously. Place emphasis on the "official" sites, as these are most likely to remain stable reference points for trainees. When creating your materials and exercises, as much as possible, avoid screen-shots, as these date quickly—otherwise, you risk spending substantial amounts of time updating outdated slides rather than concentrating on developing suitable case studies and examples relevant to your audience. Describe the essence of data that can be retrieved from a particular resource and the principles governing a tool, rather than sticking to specific releases, web interfaces, or, for example, to tables of ranked results, which are likely to differ from day to day, as new data become available in the databases. Take into account that new data may have been added to the databases you are planning to use, and hence the outputs of the queries might be different from those you planned to demonstrate. As this occurrence is actually an integral part of bioinformatics, this can be beneficial for trainees to witness—you might even want to explore such situations extensively, to convey the idea that resources and tools are dynamic.

## Rule 6: Balance Concepts with Practical Outcomes

Bioinformatics training encompasses a vast amount of learned skills. Acquiring these skills is a bit like learning to ride a bicycle, where it is best to just start pedalling, because watching others will not help you learn the process! Of course, it is important to provide trainees with the fundamental concepts and theoretical background to ensure that they can use bioinformatics tools and resources meaningfully. Nevertheless, it is a good rule to provide a balance between the theoretical/technical and contextual aspects. For example, many trainees may not value information on flat-files, relational schemas, APIs, and web services, but will be more concerned about knowing which tools and resources to use for their specific needs, and why, and how to interpret their outputs (just as the average cyclist is not interested in the internal workings of the gearbox, as long as they know how and when to shift gear!). Discuss the limitations of the methods without getting carried away by the intricacies of the algorithms or the minutiae of a tool's capabilities. Ensure that you cover not only those questions that bioinformatics approaches can answer, but also the limitations of bioinformatics, explicitly illustrating examples that cannot be answered.

Avoid long sessions of browsing around web interfaces or showing one screenshot after another. Trainees will be eager to try tools themselves and will benefit far more from a well-planned session, with adequate time allocated to an exercise or simple exploration, than from merely watching someone else explore for them. When giving a demonstration, try to get participants to follow along with you. To compensate for the likely diversity in speed and computer-ease of your audience, when possible, pair trainees of different backgrounds together and progress activities at a speed that will allow all trainees to keep pace. Once you have completed a task, confirm that everyone has achieved the result, and recapitulate the scope of the actions to reinforce the meaning and significance of the session. If you allow trainees to work by themselves on specific tasks, conclude with what you expected them to have achieved and how! Also consider providing this summary of steps and expected outcomes in an electronic/

paper version as an addendum, as trainees might want, and would certainly benefit from being able, to review the task again, on their own time. Furthermore, trainees will often be eager to share what they have learnt when they return to their work environments, so having a set of good course manuals/practical exercises is essential to enable them to do so. Absolutely avoid spending 80% of the session talking and then rushing through the last 20% of the practical aspects. Moreover, try to avoid telling trainees to finish later (on their own) whatever they did not complete, as they will probably not do so, will feel resentful because what they really wanted to do was not done and, more importantly, they will have lost the important recap and reinforcement that you can provide.

## Rule 7: Reinforce Learning with Contextual and ''Real World Experience'' Examples

Wherever possible, provide appropriate biological context: examples without relevant context lack meaning and fail to engage trainees. After introducing a new concept, allow time to put the concept immediately into action. Begin hands-on exercises with a short worked example where everyone can complete contextual learning on a common dataset. Follow this with time for further exploration: here, you might either provide a second dataset or, if relevant or practicable, invite trainees to use their own. If appropriate, illustrate examples taken from your real world research experience. For instance, outline biological problems that you tackled with bioinformatics and describe resources and tools that you adopted to solve them and to achieve your findings and how.

## Rule 8: Ensure the Methods/ Tools Have Relevance to the Trainee Experience and Scientific Research Needs

Design your materials such that the examples you provide illustrate the concepts you wish to convey and, at the same time, are relevant to the research interests of at least some of the trainees. Whenever prior information about trainees' interests is available, use it. Appreciate that a plant biologist will not have a need for human-centric examples, nor will they find them comparable. The more relevant you make the examples for the trainees, the more likely they are to retain their interest and develop their skills! Furthermore, encourage trainees to explore the tools and resources presented during the course not only with the carefully prepared examples provided, but also from the perspective of their own research interests: nothing motivates as much as the need to solve one's own problems!

The use of tools and resources from the perspective of personal research interests, will lead new users to take a fresh critical look at them. From this perspective, trainees might be able to provide a special assessment of the tools and resources introduced in the course which would be different and complementary to the one that experienced users can provide. Trainers can gain an understanding of how easy (or hard) exploring web interfaces or programmatically access and parse resources is, and specific comments on what is intuitive or not to trainees can be captured informally or formally (e.g., through surveys). In this regard, you may explain to trainees that evaluation and feedback collected during the actual training course or in a final feedback survey can aid significantly to improve bioinformatics resources.

## Rule 9: Allow for Interactivity and Provide Time for Reflection, Individual Analysis, and Exploration

Ensure interactivity and time for reflection. Provide time for trainees to acquaint themselves with the interfaces of the tools/ resources, and to understand their contents: allowing trainees to explore a tool or resource on their own tends to promote greater retention of concepts.

Schedule 10–15 minutes at the end of each module to review the presented concepts, and to stimulate questions from the trainees, who will probably have only just started processing the information.

Do not simply rely on a set of slides and step-by-step tutorials to teach concepts. Make use of flip-charts to brainstorm together, asking trainees for ideas and alternative ways to resolve particular biological questions. Group sessions like this, where trainees are encouraged to share their thoughts and views with the whole class, can help both to identify common issues and aspects to be explored, and to highlight any trainee limitations and/or mismatched expectations. Moreover, incorporating such group discussions directly into training sessions can often help to instil a greater level of understanding than when trainees are left to passively explore set examples (or to copy and paste scripts with no explanation of what these might achieve). Exploit such brainstorming sessions to demonstrate how bioinformatics tools and resources can help to address, and sometimes solve, complex problems.

Depending on the time available, include quizzes and/or problem-solving tasks and open discussion sessions in which participants can reflect on the skills they've learned and how these might be used to address questions of interest to them.

Provide trainees (perhaps in pairs or groups) with a brief set of questions prior to, and after, the training course. Questions that probe their knowledge and understanding of bioinformatics are useful both for trainers (to verify that the course has been pitched correctly and to establish what knowledge has been gained) and for trainees. Furthermore, by asking trainees to think about, and answer, a series of course-relevant questions, you ensure adequate time for concept and content digestion and reflection.

## Rule 10: Encourage Independent Thinking and Problem Solving

Finally, teach to fish rather than give fish! In other words, try to develop independent thinking rather than simply spoon-feeding trainees with slides and step-by-step tutorials: it is more important to learn how to tackle research questions with bioinformatics, and to know where/ how to search for solutions, than it is to learn about the minutiae of every available tool and resource.

## References


1. Schneider MV, Watson J, Attwood T, Rother K, Budd A, et al. (2010) Bioinformatics training: a review of challenges, actions and support requirements. Brief Bioinform 11: 544–551.
2. Schneider MV, Walter P, Blatter MC, Watson J, Brazas B, et al. (2011) Bioinformatics Training Network (BTN): a community resource for bioinformatics trainers. Brief Bioinform. In press.

Editorial

# Ten Simple Rules for Teaching Bioinformatics at the High School Level

**David Form**[1], **Fran Lewitter**[2]*

**1** Science Department, Nashoba Regional High School, Bolton, Massachusetts, United States of America, **2** Bioinformatics and Research Computing, Whitehead Institute, Cambridge, Massachusetts, United States of America

Given the availability of free, online genomic databases and tools for the analysis of biological data, it is now feasible to teach bioinformatics in the high school classroom [1]. There are a number of reasons why it is appropriate and desirable to introduce bioinformatics at the high school level. Students can engage in inquiry-based activities that involve approaching real-world problems using 21st century skills, while being tailored to high school biology frameworks. Many tools, such as 3-D protein visualization software, allow for differentiated and highly interactive instruction. The foremost reason may be that students can develop a research toolkit that they will be able to use subsequently during college and beyond.

As a high school science teacher for the past 23 years, I (DF) have had the opportunity to incorporate bioinformatics into my courses to enrich the teaching of concepts of molecular biology, human biology, genetics, and evolution, providing increased opportunities for effective differentiated instruction and individual student research. This past experience has inspired the creation of this set of Ten Simple Rules.

It is important to distinguish between curricula designed to teach the fundamentals of bioinformatics and those that utilize bioinformatics as a teaching tool. Examples of both types of successful teaching can be found in Text S1, Text S2, and Text S3.

## Rule 1: Keep It Simple

Set one, or a very few, objectives for each activity. Begin with a few, limited, straightforward goals. For example, an activity may require students to find a limited set of specific information in a GenBank file, such as the coding sequence for a gene, and print it out in FASTA format. You can link these objectives to other, more complicated, concepts in later lessons.

An activity will be more effective if extraneous information is kept to a minimum. The output provided to the students is likely to contain too much information for them to digest during one lesson. Focus on one or a few items.

## Rule 2: Familiarity: Use Activities to Explore Examples That Are Familiar to Students

Familiarity breeds relevance. Much of the information presented to students will be new to them. It will make it easier to understand new concepts or information if they are linked to something that is already familiar to them. High school students are particularly interested in topics that they can relate to their immediate personal or social lives. Choose genes, proteins, or processes that relate to disease, development, or other aspects of human physiology and behavior. Obesity, diabetes, and developmental disorders are some examples that have worked well.

## Rule 3: Link Activities to Preexisting Science Curricula

Bioinformatics exercises are more likely to be used if they are related to the curriculum that is already being taught. In a biology class, a lesson using 3-D protein models is more likely to be utilized if the proteins studied relate to concepts in the curriculum. For example, analysis of hemoglobin structure can be part of units on the circulatory system and genetics (sickle cell disease). The use of 3-D models can be used to help introduce students to structure–function relationships in proteins. Students can utilize 3-D protein models to compare the structures of proteins with very different functions, such as collagen, the estrogen receptor, and alpha amylase.

## Rule 4: Develop Activities That Build on Each Other

More complex tasks and skills can be done successfully if they are broken down into small pieces that are taught separately and then combined in a stepwise fashion. Students can focus on learning one skill or concept at a time.

## Rule 5: Use Activities to Build Skills and to Provide Information through Inquiry-Based Research

Students learn best when the work has meaning and when they are actively pursuing a goal. For example, a student who was asked to find the mRNA sequence for the gene involved in a disease that she was researching was wondering why there were several mRNA sequences for what she thought was a single gene. After an explanation of alternative transcripts and the roles of introns and exons in generating these transcripts, she was excited about her "discovery" and proceeded to explain this to her classmate/friend. She found the concept of RNA editing to be fairly easy because she actively discovered the process as part of her research.

**Citation:** Form D, Lewitter F (2011) Ten Simple Rules for Teaching Bioinformatics at the High School Level. PLoS Comput Biol 7(10): e1002243. doi:10.1371/journal.pcbi.1002243

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** October 27, 2011



**Funding:** The authors were supported by institutional funds. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

Fran Lewitter is Education Editor of *PLoS Computational Biology.*

* E-mail: lewitter@wi.mit.edu

## Rule 6: Provide Opportunities for Individualization

Students will often become more involved if they feel a sense of ownership for their work. Have individual students, or student groups, each research their own gene or protein. For example, each student in a class can be asked to identify the gene and protein associated with a unique genetic disorder. Make sure that the level of difficulty is appropriate for the level and age of the students.

## Rule 7: Address Multiple Learning Styles

Student abilities and learning styles will vary among the class. Make use of the multiple ways that information is presented. For example, the output of BLAST makes use of a colorful graphical interface, a "hit list" in chart format, and sequence alignments. Using all of these will help students to understand a BLAST output.

## Rule 8: Empower Students

Students like solving problems and discovering new information. Allow students to discover the concept or information that you want them to learn. This plays to a real strength of bioinformatics as a teaching tool. Set up activities so that students can follow up and extend their knowledge on their own, using the skills that they have developed.

## Rule 9: Model Processes Using Pen and Paper before Using the Computer

Computers can handle large amounts of data and make complex manipulation of this data in a short period of time—that's why we use them in bioinformatics. However, this can often hide the processes from the students. Have the students run through a simplified mock-up of the data analysis using pencil and paper. For example, have them compare protein sequences and come up with a "score" of relatedness before using a program, such as BLAST (through the NCBI website). Have them find and highlight appropriate data in a printed form of a BLAST readout before they analyze a BLAST readout online by themselves.

## Rule 10: Produce a Product

Have the students use the results of their activity to produce a "product" they can present to the class. If they are researching the structure and function of a protein, have them design a product that uses this protein. For example, in researching leptin they can design an obesity pill.

## Supporting Information

**Text S1 Examples of model curriculum.** Here we provide example curriculum for two types of courses for secondary school students. One is for bioinformatics activities to incorporate in an introductory biology course. The second is for a course "Models for Disease" and is offered to Accelerated/Honors level students after completing a first course in biology.
(DOC)

**Text S2 Example term project for "Models of Disease" class.** For the "Model for Disease" course, students are required to complete a term project that uses bioinformatics tools to study a disease. Here we provide an example presentation given by a student based on their term project.
(PDF)

**Text S3 Tips for developing curriculum.** The materials presented here were also presented as part of a tutorial "Teaching Bioinformatics in High School Biology Courses" held at the International Society for Computational Biology's annual meeting (ISMB) held in Boston, Massachusetts, in July of 2010.
(PDF)

## Acknowledgments

We wish to thank Curtis Huttenhower for helpful discussions.

## Reference

1. Wefer SH, Sheppard K (2008) Bioinformatics in high school biology curricula: a study of state science standards. CBE Life Sci Educ 7: 155–162.

Editorial

# Ten Simple Rules To Combine Teaching and Research

**Quentin Vicens[1], Philip E. Bourne[2]***

1 University of Colorado, Boulder, Colorado, United States of America, 2 Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, California, United States of America

The late Lindley J. Stiles famously made himself an advocate for teaching during his professorship at the University of Colorado: ''If a better world is your aim, all must agree: The best should teach'' (http://thebestshouldteach.org/). In fact, dispensing high-quality teaching and professional education is the primary goal of any university [1]. Thus, for most faculty positions in academia, teaching is a significant requirement of the job. Yet, the higher education programs offered to Ph.D. students do not necessarily incorporate any form of teaching exposure. We offer 10 simple rules that should help you to get prepared for the challenge of teaching while keeping some composure.

## Rule 1: Strictly Budget Your Time for Teaching and for Doing Research

This rule may seem straightforward, but respecting it actually requires more discipline and skill than it first appears to. The key is to set aside time for both teaching and research from the beginning, with a well-marked separation (e.g., mornings will be devoted to course preparation, afternoons to experiments and manuscript writing). Firmly stick to this agenda, particularly if this is your first time teaching. Failure to do so would eventually affect the quality of your teaching or the progress of your research (or both). Over time, you will become more skilled at jumping from one commitment to the other, and therefore allowing the boundaries to fluctuate somewhat. Avoid underestimating the time necessary to fulfill teaching-related obligations (e.g., office hours, test preparation, grading, etc.) by consulting with your colleagues.

## Rule 2: Set Specific Teaching and Research Goals

In order not to have one occupation overpower the other one—which would transgress Rule #1—it is a good idea to decide on specific aims for each enterprise. Compile a list of reasonable but specific long-term goals (for the month or the semester) and short-term ones (for the week) for both your teaching (e.g., finish Chapter 3 by Nov. 1; this week propose a discussion to engage students to brainstorm about the risks of GMOs) and your research (e.g., finish experiments for this project and start writing before Easter; this week do the control for my primer binding assay). Make sure you achieve them. If you don't—this is likely to happen at first—ask yourself how legitimate your reason is. Then review and adjust the goals accordingly.

## Rule 3: ''Don't Reinvent the Wheel''

We borrowed the title for this rule from excellent suggestions on *How To Prepare New Courses While Keeping Your Sanity* [2]. Most likely, you will not be the first one ever to teach a particular topic. So get in touch with the colleagues in your department who have taught the class you are going to teach, or who teach similar topics. You can also use your network and contact former colleagues or friends at other institutions. They will usually be happy to share their course material, and along the way you might also glean precious tips from their teaching experience (e.g., a list of do's and don'ts on how to approach a notoriously difficult topic). You will also learn a lot from sitting in one of their classes and watching how they handle their topic and their students. Here are more examples of precious time-savers:

(1) Choose a textbook that is accompanied by rich online resources such as annotated figures, pre-made PowerPoint slides, animations, and videos. Students will thank you for showing movies, for example, as they often are a better option to break down complex mechanisms or sequences of events into distinct steps.

(2) Administer a Web site for your course. Many universities and some textbooks now offer you the possibility of hosting a Web site with course-related materials, including automatically graded assessments. See, for example, the CULearn suite used at the University of Colorado (http://www.colorado.edu/its/culearn/), or more general automatic grading tools presented at http://ctl.stanford.edu/Tomprof/postings/227.html.

(3) Gather a solid team of motivated teaching or learning assistants, who will both serve as an intermediary between you and your students and help you grade. In short, don't be afraid to ask for help!

## Rule 4: Don't Try To Explain Everything

Class time should be spent guiding students to create their own explanation of the material and to develop cognitive abilities that will help them become critical thinkers. In other words, you don't want to present all aspects related to a certain topic or to lay out all the explanations for them. Thus, an effective way to teach is to get students to learn by transformative learning: beyond memorizing and comprehending basic concepts, they will learn to reflect on what they learn and how they learn it (see, for example, http://en.wikipedia.org/wiki/Transformative_learning and references within). Such teaching practices require that a significant part of the learning process happens outside the classroom, through reading assignments, homework, writing essays, etc. So make sure you budget time to organize these, as specified

**Citation:** Vicens Q, Bourne PE (2009) Ten Simple Rules To Combine Teaching and Research. PLoS Comput Biol 5(4): e1000358. doi:10.1371/journal.pcbi.1000358

**Published** April 24, 2009



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: bourne@sdsc.edu

Philip E. Bourne is the Editor-in-Chief of *PLoS Computational Biology*.

in Rule #2. Remember that in the end this will be a win–win situation: you will save time by not having to fit everything into your class time, and students will learn how to find answers through their own thinking.

## Rule 5: "Be Shameless in Bringing Your Research Interests into Your Teaching"

This is yet another great time-saver, and this rule title is actually from *Confessions about Stress and Time: Thoughts for Faculty* (available at http://www.colorado.edu/ftep/publications/confessions.html). Students want to know how what you teach relates to the world around them. They also like to know what is happening in science right now, so this is where you can feed in some of your research interests (for some examples of how researchers around the world have been bringing their research into the classroom, refer to the special section of the July 6, 2007, issue of the magazine *Science* entitled *The World of Undergraduate Education* [3]. Students will welcome such connections, especially in an introductory course or in a course for non-majors. Additionally, they will feel the passion that makes you love being a scientist. On your end, you might find that preparing course materials will be easier (because you are already a master of that topic), and you might learn to be more comfortable at presenting your research in layman's terms.

## Rule 6: Get the Most in Career Advancement from Bringing Your Research into Your Teaching

As a sort of followup to Rule #5, presenting your research in class could bring you a solid return on your investment. For example, teaching gives you exposure; talking about your research may help you recruit motivated students in your lab, which will help you advance your research, possibly by taking it in original directions. In parallel, you could also use your research to design a novel course and possibly evaluate student learning in a fashion that would make for a publication in a science education journal. Another option would be to write or edit a book, or to contribute a chapter in someone else's book that you would eventually give as a reading assignment in your class. Conversely, there is wisdom in crowds. Consider having students review aspects of your research that fit the course and get feedback. You will be surprised at what useful information can come from students critiquing a new manuscript or proposing new experiments.

## Rule 7: Compromise, Compromise, Compromise

A significant part of the compromise once you accept a joint research/teaching commitment is to realize that your list of "things that in principle you would like to do but won't have time to do" will get longer. Maybe you would like to personally respond to all the students who e-mail you about any problem they may have, but, realistically, such things can't happen. Instead, a solution would be to send some general feedback in answer to the common queries and to write occasional brief personal responses. As you get more skilled at combining research and teaching, you will be able to progressively bring back activities such as scanning the most recent scientific literature and attending seminars and lectures more often. But remember to accept that no matter how skilled you are at budgeting your time for teaching and research, you will still face the conflicting demands of both, and you will have to keep compromising. In the end, compromising will sometimes imply learning to say no when pondering about taking on a novel and exciting assignment that would unequivocally conflict with your current research/teaching agenda.

## Rule 8: Balance Administrative Duties with Your Teaching and Research Workload

Your responsibility as a teacher and as a researcher is to be as productive as you can be in these two areas, at the same time. This is what your colleagues and the faculty board will expect from you when evaluating you for tenure, for example. Doing service within your community (for example by sitting on committee meetings, or by being part of a local scientific club) counts as well, but not as much. In consequence, turning down yet another offer to organize a series of seminars, or to edit the newsletter of your department, is legitimate if it cuts into your productivity. Similarly, keep your ability to career advance in mind when considering taking on another teaching assignment.

## Rule 9: Start Teaching Early in Your Career

This will be the best way to get exposed to some of the difficulties mentioned in the other Rules sooner rather than later. You can see this as an opportunity to learn how to add on various responsibilities in a gradual rather than an immediate manner (e.g., when "jumping" from a post-doc to a faculty position at a university). Many options are available to teach at the graduate level (e.g., by becoming a teaching/learning assistant), as well as at the post-graduate level (e.g., by teaching part-time on campus or at a local school while doing your post-doc). You may need to be proactive about looking for such opportunities, but an increasing number of universities and institutions are developing programs that formally offer teaching experience to graduate students and post-docs [4,5].

## Rule 10: Budget Time for Yourself, Too

A lot of stress can build up from a constant shuttle between teaching demands and research occupations. In order to be able to evacuate some of that tension, it is a good idea to hide some time for yourself that you will spend with your family, or to do your hobby, to exercise, to travel, etc. An unfulfilling personal life is incompatible with successful teaching and research careers. Consequently, don't forget to spend some energy learning how to balance both areas.

Finally, keep in mind that your experience can make for a valuable contribution to the scientific community, for example, in the form of a report on your efforts in science education, or by posting comments to this Editorial!


## Acknowledgments

We thank Joe Ma, Clayton Lewis, and Jamie Williamson for careful reading of the manuscript.


## References


1. Editorial (2007) Those who can teach, should. Nat Chem Biol 3: 737.
2. Brent R, Felder RM (2007) Random thoughts: How to prepare new courses while keeping your sanity. Chem Engr Education 41: 121–122. [Reprinted in a posting by Rick Reis on the Tomorrow's Professor mailing list at the Stanford University Center for Teaching and Learning at http://ctl.stanford.edu/Tomprof/postings/800.html].
3. Mervis J (2007) Special section—The world of undergraduate education. Science 317(5834): 63–81.
4. Coppola BP, Banaszak Holl MM, Karbstein K (2007) Closing the gap between interdisciplinary research and disciplinary teaching. ACS Chem Biol 2: 518–520.
5. Tahmassebi DC, Williamson JR (2007) Balancing teaching and research in obtaining a faculty position at a predominantly undergraduate institution. ACS Chem Biol 2: 521–524.

OPEN ACCESS Freely available online
PLoS COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Graduate Students

Jenny Gu, Philip E. Bourne*

Choosing to go to graduate school is a major life decision. Whether you have already made that decision or are about to, now it is time to consider how best to be a successful graduate student. Here are some thoughts from someone who holds these memories fresh in her mind (JG) and from someone who has had a whole career to reflect back on the decisions made in graduate school, both good and bad (PEB). These thoughts taken together, from former student and mentor, represent experiences spanning some 25 or more years. For ease, these experiences are presented as ten simple rules, in approximate order of priority as defined by a number of graduate students we have consulted here in the US; but we hope the rules are more globally applicable, even though length, method of evaluation, and institutional structure of graduate education varies widely. These rules are intended as a companion to earlier editorials covering other areas of professional development [1–7].

## Rule 1: Let Passion Be the Driving Force of Your Success

As with so many other things in life, your heart and then your head should dictate what thesis project makes sense to embark on. Doing your best work requires that you are passionate about what you are doing. Graduate school is an investment of up to a seven-year commitment, a significant chunk of your life. Use the time wisely. The educational system provides a variety of failsafe mechanisms depending on the part of the world where you study. Laboratory rotations and other forms of apprenticeship should not be overlooked, for they are opportunities to test the waters and measure your passion in a given subject area. It is also a chance to test your aptitude for research. Take advantage of it! Research is very different from simply taking courses. If you do not feel excited about doing research and the project selected, do not do it; reevaluate your career decisions.

## Rule 2: Select the Right Mentor, Project, and Laboratory

Finding the right mentor can be hard since it is not always possible to know the kind of mentoring that is going to work best for you until you actually start doing research. Some of us like to work independently, others like significant feedback and supervision. Talk to other students in the laboratory and get their impressions of how the principle investigator's mentoring works for them. In a large laboratory, chances are you will get less direct mentoring from the principle investigator. In that case, other senior scientists in the laboratory become important. What mentoring are they likely to offer? Judge, as best you can, if the overall environment will work for you. A key element is the standing of your mentor in his or her scientific field. When you graduate, the laboratory you graduate from is going to play a role in determining what opportunities exist for your postdoctoral work, either in academia, industry, or other sectors. Your proposed mentor should be very enthusiastic about the project you discuss. If he or she is not, you have the wrong mentor and/or project. At the same time, beware that such enthusiasm, however senior the mentor, may be misplaced as far as your interests are concerned. Gauge the novelty of the research project and potential for high-quality publications by doing your own background check through reading previously published research and talking to other scientists in related areas. Also consider if the project can be reasonably completed in the allocated time for graduation. To propel your career, you want to come out of a higher degree as a recognized individual having made a significant scientific contribution. Thus, it is absolutely critical that you do take the time to find the project and mentor that is going to fulfill this goal.

## Rule 3: Independent Thinking Is a Mark of a True Scientist

Regardless of your initial work habits and how much you depend on your mentor (Rule 2), eventually you will have to be more independent than when you started graduate school. The earlier you start on that path to independence the better. Independence will play a critical part in your career as an innovative scientist. As much as possible define your own research project with a view to make a significant and unique scientific contribution.

## Rule 4: Remember, Life Is All about Balance

Take the time to meet your own needs. Graduate school is highly demanding, both mentally and physically. Your health comes first, spend the time being healthy or else you might find yourself spending more time being sick. Hard work should be balanced with other activities that you enjoy and give you a break. These activities can often become important in your future scientific career. Collaborations sometimes start not because of a shared scientific interest initially, but because you share the same hobby or other interest.

## Rule 5: Think Ahead and Develop Your Professional Career Early

There are two parts to this. The first part relates to professional development. Being a successful scientist is more involved than just doing good science. You need to be able to write good papers, submit compelling scholarship and grant applications, make powerful

**Citation:** Gu J, Bourne PE (2007) Ten simple rules for graduate students. PLoS Comput Biol 3(11): e229. 10.1371/journal.pcbi.0030229



Jenny Gu is with the University of Texas, Medical Branch, Galveston, Texas, United States of America. Philip E. Bourne is with the University of California San Diego, La Jolla, California, United States of America.

* To whom correspondence should be addressed. E-mail: bourne@sdsc.edu

presentations, and communicate and collaborate with other researchers. The other Ten Simple Rules editorials are a start here [1–7], but you need to work on developing these skills at the same time as you work on your thesis. The second part involves using these emergent skills to figure out what to do with the higher postgraduate degree. Do not wait until you graduate to take the next step. Have a position and a fellowship, if possible, lined up ahead of time.

### Rule 6: Remain Focused on Your Hypothesis While Avoiding Being Held Back

Formulation of the hypothesis is the first thing you'll learn in Science 101, and yet somehow it seems to get occasionally thrown out the window. When you find yourself lost in the details of your research, take a step back and remind yourself of the big picture. Revaluate your hypothesis from time to time to see if it still makes sense, because you may find yourself needing a new one. Always keep this in mind in discussions with your mentor. As you have these discussions, remember you are cheap labor, and, if you are a good student, a source of success to your mentor. The temptation is that your mentor will want to keep you around as long as possible. Define the scope of your project early with your mentor and agree that this is what you will attempt to complete in order to receive the degree. A career awaits you beyond the laboratory of your graduate student days. Do not prolong moving on to new challenges.

### Rule 7: Address Problems Earlier Rather Than Later

If graduate school wasn't quite what you thought it would be, be it scientifically or otherwise, find out what your options are to address the problem. Discuss these problems with your mentors. A good mentor is there not just to guide you scientifically, but also in your personal development. Remember, they have been there themselves and have likely seen similar issues with earlier students. Take time off to reflect on your future if this is needed. A good mentor will understand that you come first.

### Rule 8: Share Your Scientific Success with the World

Being recognized by your peers as someone who does good science is important both within your institution, nationally, and internationally. When opportunities arise to give seminars and presentations to other groups, take them. Before starting with a mentor, come to an agreement as to when and what meetings you can attend locally and globally. Scientific meetings are a fun and fruitful venue for exchange. Be sure to venture beyond the comfort zone of familiar faces, because it is important to meet other colleagues in your field. These people may become your future collaborators, friends, advocates, and employers.

### Rule 9: Build Confidence and a Thick Skin

As you pave the road to scientific fame with Rule 8, expect your work to be criticized and scoffed at, for that is part of the scientific process of challenging new ideas. The best way to build self-confidence for these otherwise defensive moments is to be prepared and to present your work clearly with a confident display of your expansive knowledgebase of the relevant related work. Do not be intimidated by big names who question your work; counter knowledge with knowledge. Another reason to have a thick skin is that the path to success will not be without setbacks—setbacks such as experiments that fail, and experiments that succeed but do not yield a useful result causing you to have wasted significant time. Undergraduate training is usually much more structured and does not prepare you for such setbacks. Learn as much as you can from these situations both about the science and yourself and move on.

### Rule 10: Help Select and Subsequently Engage Your Thesis Committee

This rule depends somewhat on how your institution is structured. Some institutions do not convene a thesis committee until near the end of your work. For those institutions that require a thesis committee to be convened early, talk with your mentor and be involved in the selection process. The committee is there to work for you as secondary mentors. Consider people whose own research experience will be valuable to you or who have a reputation for ongoing mentoring in all areas of professional development. Make a point of talking to members of the committee from time to time and keep them abreast of what you are doing. On occasion, you and your primary mentor may have disagreements; committee members can be invaluable here. ■


## Acknowledgments

Thanks to Kristine Briedis, Jo-Lan Chung, Ruben Valas, and Song Yang, current and former students in the Bourne Laboratory, and members of the Bioinformatics and Chemistry/Biochemistry Programs at the University of California San Diego, for their insightful comments on these rules. As always, we invite you to comment, either formally through the journal, via blogs and list servers, and to the authors directly.

**Funding.** The authors received no specific funding for this article.

**Competing interests.** The authors have declared that no competing interests exist.


## References


1. Bourne PE (2005) Ten simple rules for getting published. PLoS Comp Biol 1: e57. doi:10.1371/journal.pcbi.0010057
2. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comp Biol 2: e12. doi:10.1371/journal.pcbi.0020012
3. Bourne PE, Korngreen A (2006) Ten simple rules for reviewers. PLoS Comp Biol 2: e110. doi:10.1371/journal.pcbi.0020110
4. Bourne PE, Friedberg I (2006) Ten simple rules for selecting a postdoctoral fellowship. PLoS Comp Biol 2: e121. doi:10.1371/journal.pcbi.0020121
5. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comp Biol 3: e44. doi:10.1371/journal.pcbi.0030044
6. Bourne PE (2007) Ten simple rules for making good oral presentations. PLoS Comp Biol 3: e77. doi:10.1371/journal.pcbi.0030077
7. Erren TC, Bourne PE (2007) Ten simple rules for a good poster presentation. PLoS Comp Biol 3: e102. doi:10.1371/journal.pcbi.0030102

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for organizing a non–real-time web conference

**Ana Arnal**[1,2], **Irene Epifanio**[1,2,3]*, **Pablo Gregori**[1,2], **Vicente Martínez**[1,2]

**1** Departament de Matemàtiques, Universitat Jaume I, Castellón de la Plana, Spain, **2** IMAC (Institut de Matemàtiques i Aplicacions de Castelló), Universitat Jaume I, Castellón de la Plana, Spain, **3** IF (Institut d'Estudis Feministes i de Génere), Universitat Jaume I, Castellón de la Plana, Spain

* epifanio@uji.es





**Citation:** Arnal A, Epifanio I, Gregori P, Martínez V (2020) Ten Simple Rules for organizing a non–real-time web conference. PLoS Comput Biol 16(3): e1007667. https://doi.org/10.1371/journal.pcbi.1007667

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** March 26, 2020



**Funding:** This work has been partially supported by the Educational Support Unit (USE-Unidad de Soporte Educativo) of Universitat Jaume I of Castellón (grants 3595/18, 3596/18, 3598/18, 3874/19). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

The present work describes the 100% virtual ATIDES (Avances en Tecnologías, Innovación y Desafíos de la Educación Superior) conference that was held between October 15 and 31, 2018, sponsored by Universitat Jaume I (UJI), Spain. Online conferences like this have been the subject of some controversy in the field of education over the last decade. Indeed, we have found a few texts that are against them. One of these is [1], whose authors claim that "interaction is not enough" to ensure efficient simulation of face-to-face contact. However, the Canadian academic community (for instance, the Centre for Distance Education at Athabasca University) is a strong advocate of online conferences (see [2,3]). Among other advantages, this kind of conference is "family-friendly," i.e., they break barriers for researchers with family obligations [4], in particular many women [5]. In addition, these conferences overcome the drawback of parallel sessions at face-to-face conferences, at which participants must choose certain talks and miss others. Anderson and Anderson [6] even put forward environmental and economic arguments: "Transportation is a major contributor of carbon dioxide (CO2) emissions." On the other hand, Abdullah [7] and Kear, Chetwynd, and Jefferis [8] look at the matter from another point of view that is also important: Social presence at online conferences.

Gichora and colleagues [9] propose 10 rules for organizing a virtual conference, but they only focus on real-time virtual conferences. Real-time virtual conferences emulate the structure of well-known on-site conferences, with the use of video-conferencing tools for connecting the participants in a virtual meeting. We, on the other hand, focus on non–real-time web conferences. In this kind of conference, a web platform connects the participants, but the event does not take place live. Communication can be made through posts on a forum, which can sometimes take hours or even days.

Non–real-time web conferences share many of the advantages of real-time virtual conferences, such as being family friendly and allowing the involvement of participants with a low budget for traveling, which makes this kind of conference more participative and inclusive. However, non–real-time web conferences have other advantages that real-time web conferences do not. Asynchronous interaction allows the attendees to have more time for reflection before they put forward their questions and comments, and, likewise, the speaker has enough time to prepare an appropriate answer. This makes up for the lack of face-to-face social interaction in this type of conference. In fact, numerous interactions occur via forums and are, in many cases, numerous and more informal and flexible than at classical scientific meetings [10]. Forums give rise to stimulating discussions like those that can occur at unconferences [11]. Furthermore, this asynchrony allows participants from all over the world to take part without worrying about time zone differences.

Another advantage is that it is not so greatly dependent on technology, unlike real-time web conferences: video-conferencing software is not necessary nor is a stable internet connection. In fact, we do not have to worry about the bandwidth like we do at real-time conferences. So, people in places that might not have very good or reliable internet connectivity can also participate.

Therefore, the cost of organizing a non–real-time web conferences is lower than for other scientific events. This means that the registration fees can be made very cheap (the registration fee for ATIDES 2018 was €35). This makes these conferences more affordable for everyone.

Furthermore, these conferences are more accessible since communications in non–real-time web conferences are text based. Although the conference can have an official language, language is not a barrier since Google Chrome can translate the text, so the content is accessible in any language without having to hire translation and interpreting services. Not only that, the content is also accessible for people with auditory disabilities.

In this paper, we describe the most notable data that have arisen from the event. Furthermore, we detail the steps taken previously during the planning stage as well as an evaluation of the results obtained. This conference was developed 100% virtually from beginning to end: The call for papers, the paper-selection process, the preparation and announcement of the proceedings, the process of presentations, discussions in forums, and issuance of the corresponding attendance and authorship certificates.

The following Ten Simple Rules are the result of the experience obtained from organizing the two editions of the 100% virtual ATIDES conference.

## Rule 1: Set up an organizing and scientific committee that is engaged with information and communications technologies

Effective committees are essential for the success of every kind of conference. In particular, a web conference also requires its organizers to have computer skills or, at least, have some knowledge of the computer tools that exist nowadays. As we will explain here, we had to handle web-hosting platforms and conference-management tools, and moreover, we created our own website. Therefore, an experienced organizing committee (supported by an expert information technology [IT] team, see Rule 3) makes this task easier. Regarding the scientific committee, its members should, at least, be aware of and somewhat familiar with information and communications technologies (ICTs), since they will have to interact by computer during the paper-selection process.

In our case, the ATIDES 2018 organizing committee was formed by researchers and teachers who were committed to the theme of the conference and willing to work as a team with ICT tools. They were also part of the scientific committee to ensure efficient coordination. The scientific committee consisted of 39 renowned researchers in teaching innovation from some of the universities listed in Table 1.

## Rule 2: Correct timing of the conference

Exchanges in a non–real-time conference generally take much longer than in real-time conferences. This is particularly the case when participants belong to different time zones. Therefore, it is strongly recommended to establish a length of at least one or, preferably, two weeks for the whole conference. Moreover, it is very important to choose the right dates on which to hold the conference. Overlaps with similar conferences should be avoided to obtain maximum attention from potential participants. It is also necessary to consider periods that are not occupied by other activities (lectures, holidays, etc.). Nevertheless, even in the worst case, a non–

**Table 1. Acronyms of participants' institutions.**

| | |
|---|---|
| CNCIVIRTUAL: Universidad Centro Nacional de Capacitación Intensiva, Mexico. | ITSON: Instituto Tecnológico de Sonora, Mexico. |
| UAB: Universitat Autònoma de Barcelona, Spain. | UACO: Universidad Nacional de la Patagonia Austral, Argentina. |
| UANL: Universidad Autónoma de Nuevo León, México. | UAUSTRAL: Universidad Austral, Argentina. |
| UCAVILA: Universidad Católica de Ávila, Spain. | UCLM: Universidad de Castilla La Mancha, Spain. |
| UCM: Universidad Complutense de Madrid, Spain. | UCO: Universidad de Córdoba, Spain. |
| UCOMILLAS: Universidad de Comillas, Spain. | UHU: Universidad de Huelva, Spain. |
| UJI: Universitat Jaume I, Spain. | ULPGC: Universidad de Las Palmas de Gran Canaria, Spain. |
| UNAHUR: Universidad Nacional de Hurlingham, Argentina. | UNED: Universidad Nacional de Educación a Distancia, Spain. |
| UNEX: Universidad de Extremadura, Spain. | UNILEON: Universidad de León, Spain. |
| UNIOVI: Universidad de Oviedo, Spain. | UNIRIOJA: Universidad de La Rioja, Spain. |
| UNIZAR: Universidad de Zaragoza, Spain. | UNLP: Universidad Nacional de La Plata, Argentina. |
| UPM: Universidad Politécnica de Madrid, Spain. | UPOLI: Universidad Politécnica de Nicaragua, Nicaragua. |
| UPV: Universitat Politècnica de València Spain. | URJC: Universidad Rey Juan Carlos, Spain. |
| US: Universidad de Sevilla, Spain. | USAL: Universidad de Salamanca, Spain. |
| USFQ: Universidad San Francisco de Quito, Ecuador. | UTPL: Universidad Técnica Particular de Loja, Ecuador. |
| UV: Universitat de València, Spain. | UVG: Universidad del Valle de Guatemala, Guatemala. |
| UVIC: Universitat de Vic, Spain. | UVIGO: Universidad de Vigo, Spain. |

https://doi.org/10.1371/journal.pcbi.1007667.t001

real-time conference can fit in well with such other events thanks to the choice of a sufficiently long period.

The ATIDES conference took place during the last two weeks of October 2018, and it included participants mainly from Europe and South America, with a time difference of around eight hours. One week earlier, the inaugural speech was filmed. The institutional and academic authorities of UJI gave strong support for the event and showed a positive and warm attitude towards the participants.

At the very beginning of the conference, all attendees received a welcome message with a link to the virtual conference space. This message, which was sent to the general forum of the conference room, encouraged the participants to view the inaugural speech.

## Rule 3: Have an expert IT team to assist with web hosting and digital tasks

Efficient management of the ICT tools used is essential when carrying out an event like this. It is therefore necessary to have a team of expert technicians in this field who can help the organizers solve any technical problems that may occur as well as provide advice on the design and management of the virtual environment to be used. Online communication before, during, and after the conference (even in a non–real-time event) is the basis for effective interaction between all participants (organizers, speakers, and attendees), and this depends largely on the available computer support.

At the ATIDES 2018 conference, our IT team created the virtual conference space and had to react after a computer-server breakdown occurred during the course of the conference.

## Rule 4: Identify a good, visible hosting platform

Among the plethora of free and paid hosting services available, a clear and friendly website is absolutely necessary to encourage participation and increase the visibility of the conference. It is advisable to use a short and easy-to-remember web domain. The expert IT team can help with the best choice.

In our case, for the sake of greater freedom, affordability, and an elegant front end, we used a third-party free web design (https://www.weebly.com) for our first contact website.

## Rule 5: Develop the website and the virtual conference space

The website must contain all the usual key elements of a conference, such as "general information," "committees," "topics," "registration," and "access to the virtual conference space". In addition to the list of specific topics, it should also include keywords such as "web conference" and "asynchronous" (or "non–real-time") in order to ensure it is placed among the top internet search results of people seeking this kind of events.

Regarding the virtual conference space, it must be hosted on servers managed by the IT team. The IT team is responsible for installing, customizing, and maintaining a solid learning management system (LMS) program for managing the development of the conference. This software must be flexible and open, have a friendly interface, and facilitate asynchronous communication (interaction between participants can be promoted through open-forum spaces) and, optionally, synchronous communication (we offer a non–real-time event where real-time interaction is also welcome but not compulsory). The mature open-source Moodle or the more recent Blackboard Open LMS are two important examples. The constant availability of the servers is important (since authors and attendees may connect at any time of the day) but not as critical as for real-time web conferences. In the event of malfunction during the conference, the IT team must react quickly to restore connectivity.

We used our own space, which had been structured using a Moodle course to facilitate the development of the conference with a minimum hierarchical structure to organize participation in discussions.

## Rule 6: Have a good publicity campaign

Although publicity is important for any conference or event in general, in this case it can be considered a very important task to which we must pay special attention. The mechanism and development of non–real-time online conferences should be explained in the publicity campaign. In fact, a frequently asked questions (FAQ) section on the website is recommended.

Because our contact is exclusively online, the conference must be announced well in advance, and periodic calls should be made to encourage participation as speakers. The amount of notices and news should be greater than for a face-to-face conference. In this regard, it is important not to tire people with excessive information if they are not interested; they should be given the option to unsubscribe from the conference mailings.

Perhaps the place or the organizing institutions are little known, or, even if they are well known, it would offer added value to raise their profile and highlight the most attractive aspects.

Adequate diffusion of the accepted papers can help encourage participation. For example, indexing of the publication in the Conference Proceedings Citation Index by Clarivate Analytics could be requested.

In addition, in the call for papers it is important to be very clear regarding all details related to the format of the presentations, while submission deadlines must be perfectly explained on the website. As in face-to-face conferences, a call for papers was created, containing the purpose of the conference and the submission guidelines, together with important dates such as the submission deadline and the decision notification date.

## Rule 7: Specific actions for a virtual paper-selection process

Thirty years ago, works were sent by postal mail to be reviewed before being presented at a conference. Twenty years ago, these works were sent by email. But over the last decade, the use of web platforms for submission and peer review has been gaining popularity (see [12]). Thus, several conference-management tools have appeared, two of the most popular being the EasyChair conference system (see [13]) and the HotCRP conference-management system (see [14]). For more details, readers who are interested may consult Kanav, Lammich, and Popescu [15]. We cite, as alternative free platforms, ConfTool [16] and OpenConf [17]. Parallel to this, there has also been an increase in concern about verifying documentation submitted online in a confidential manner. Thus, for this new scenario the use of information security applications is necessary (see [18]). These tools are not exclusive to online conferences, but it is useful to mention them here since their use is necessary in virtual conferences.

In order to carry out the submission and review process considering the above requirements, we used the free version of the aforementioned EasyChair platform. This option allows us to handle all submissions with no restriction on the total size of uploaded files, except for a 20 Mb limit on the size of each individual file uploaded.

The platform allows us to set the parameters for the review process (e.g., reviewer assignations, reviewer forms for assessing papers, dates for opening the submission process, etc). The scientific committee, which includes all the reviewers, was entered on the platform, and invitation emails were automatically sent to all of them. Once the authors had submitted their work, two referees reviewed them. The referees assessed the contributions, and a subcommittee of the scientific committee made a decision and provided their comments and scores. The possible outcomes were rejected, resubmit with major changes, resubmit with minor changes, and accepted. The platform allows personalized emails to be sent to the authors or corresponding authors with the decision and comments from the reviewers. Those with the "resubmission" decision were given two weeks to make the changes indicated by the referees, and the subcommittee of the scientific committee then checked that those changes had been made. The format of the final accepted version was also reviewed in the EasyChair platform.

Although the scientific committee members are not required to be experts on ICTs, it is essential that they are active and willing to review the proposed papers in the aforementioned online platforms. In online conferences, papers can be more extensive than in face-to-face conferences. This is positive because papers can be more complete and detailed, but, on the other hand, the reviewers (in our case, the members of the scientific committee) have to make an extra effort.

A selection of the accepted papers for ATIDES 2018 was compiled in an electronic book with an international standard book number (ISBN). After a final review of the format of the selected publications, the proceedings book was edited with the help of our university editorial team. Moreover, although the papers were written in Spanish, we required the title, abstract, and keywords to be included in English too, since these proceedings were to be submitted for indexing in the Conference Proceedings Citation Index (Web of Science, Clarivate Analytics; see https://clarivate.com).

## Rule 8: Take actions to ensure the smooth running of the conference

Before starting the conference sessions, once the virtual conference environment is created, several necessary actions should be carried out to ensure that the conference runs smoothly.

Most of all, it is important to create good quality and up-to-date content, both on the website and in the Moodle environment, to engage and continuously provide access to the audience.

We recommend regular communication with speakers whose contributions have been accepted and periodic announcements about the conference on general news forums and social media.

Although it is important to maintain constant communication to keep the participants interested in all kinds of conferences, in this case it is essential. In order to ensure a fluent participation of the attendees and authors and their exchanges and to avoid misunderstandings in discussions, we recommend the following actions be taken in the virtual conference space:

- Structure the conference space in clear blocks under the name of each topic within the conference.
- Create a separate forum for each contribution, named with the title of the paper and containing, as an attachment, the PDF document of the full text of the paper.
- Subscribe each author to his/her corresponding forum, and encourage him/her to post an introductory message on the opening day of the conference.
- Let the attendees choose their subscription to forums. Initially, they should not be subscribed to any forums, or they will receive many unwanted email notifications.
- Automatically subscribe each participant to a forum when he/she has posted a comment. In this way, he/she will receive an email notification when an author or other attendee posts an answer to his/her comment.
- Condition the certification of attendance to participation in a prescribed number of forum and/or posts.
- Condition the certification of authorship to authors of contributions whose forums have all participants' comments answered (by any of the authors).

One way to maintain interest and encourage participation is to propose the recording of short videos where the authors present their talks (following the rules explained in Lortie [19]). In the ATIDES conference, the video presentation of each paper had to have an image quality between 480 and 720 pixels and be accessible on YouTube. The PechaKucha [20] format is based on the fact that you have limited time to convey the fundamental idea of a proposal, since the recipient soon loses attention at an event where many speakers are involved. The idea is simple: 20 slides lasting 20 seconds each, i.e., a total duration of 6 minutes and 40 seconds. The authors who opted for this modality achieved a specific participation certificate.

## Rule 9: Make the content accessible in different languages

Many scientific events have only one official language, usually English (although in ATIDES 2018 the official language was Spanish). Hiring professional interpreters for simultaneous interpreting is not so common at scientific events. However, in non–real-time web conferences, machine translation can be exploited to increase the number of participants with different languages. The content of the website, papers, and forums is written, so the translation tool in Google Chrome can break down language barriers with a single click. Although Google's take on an online interpreter will never be as good as a human translator, it can be a very good starting point.

As regards the videos of PechaKucha presentations, since they are on YouTube, its auto-caption feature can be used. Although YouTube's automatic captions are not perfect and gender and dialect bias can be found [21], they can help to overcome language barriers.

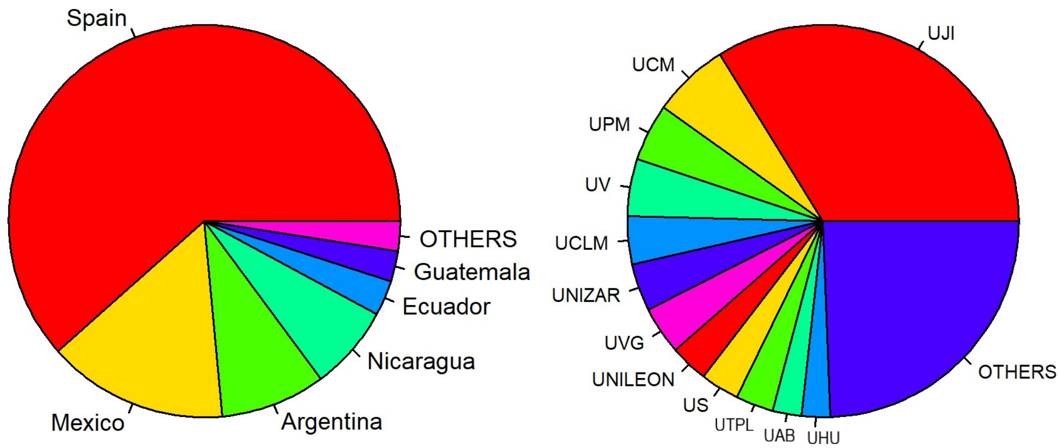



**Fig 1. Left: Participants by country.** OTHERS include participants from Venezuela, Brazil, Denmark, and South Africa. Right: Participants by institution. OTHERS includes UACO, UANL, UPOLI, UVIC, ITSON, UAUSTRAL, UCAVILA, UCO, UCOMILLAS, ULPGC, UNAHUR, UNED, UNEX, UNIOVI, UNIRIOJA, UNLP, UPCT, UPV, URJC, USAL, USFQ, and UVIGO. ITSON, Instituto Tecnológico de Sonora; UAB, Universitat Autònoma de Barcelona; UACO, Universidad Nacional de la Patagonia Austral; UANL, Universidad Autónoma de Nuevo León; UAUSTRAL, Universidad Austral; UCAVILA, Universidad Católica de Ávila; UCLM, Universidad de Castilla La Mancha; UCM, Universidad Complutense de Madrid; UCO, Universidad de Córdoba; UCOMILLAS, Universidad de Comillas; UHU, Universidad de Huelva; UJI, Universitat Jaume I; ULPGC, Universidad de Las Palmas de Gran Canaria; UNAHUR, Universidad Nacional de Hurlingham; UNED, Universidad Nacional de Educación a Distancia; UNEX, Universidad de Extremadura; UNILEON, Universidad de León; UNIOVI, Universidad de Oviedo; UNIRIOJA, Universidad de La Rioja; UNLP, Universidad Nacional de La Plata; UNIZAR, Universidad de Zaragoza; UPCT, Universidad Politécnica de Cartagena; UPM, Universidad Politécnica de Madrid; UPOLI, Universidad Politécnica de Nicaragua; UPV, Universitat Politècnica de València; URJC, Universidad Rey Juan Carlos; US, Universidad de Sevilla; USAL, Universidad de Salamanca; USFQ, Universidad San Francisco de Quito; UTPL, Universidad Técnica Particular de Loja; UV, Universitat de València; UVG, Universidad del Valle de Guatemala; UVIC, Universitat de Vic; UVIGO, Universidad de Vigo.

https://doi.org/10.1371/journal.pcbi.1007667.g001

Furthermore, it is possible to ensure website accessibility, for people with disabilities, people with slow internet connection, or users of different devices (e.g., from desktop or laptop computers to smartphones).

## Rule 10: Analyze data and obtain feedback to learn valuable lessons

All experiences must be evaluated to objectify reality and learn lessons that can be applied to future actions, which should be designed taking the available evidence into account. It is important to know and analyze the degree to which the initial objectives have been achieved so that we can design priorities and objectives that are consistent with reality. In our case, the indicators considered were the number of participants, the number of interventions in debates, the ease of handling the virtual environment, and the opinions of the actors involved during all stages of the conference (technicians, committees, participants, and speakers). Based on this evaluation, we may decide to continue with, modify, or discard the activity.

## Lessons learned from the ATIDES 2018 conference

Our non–real-time web conference, ATIDES 2018, included 242 participants and 56 papers signed by 110 authors. Fifteen of them gave a video presentation of their talk. While 80% of the papers led to at least 19 interventions, 50% of them exceeded 33 exchanges, and one of them even obtained 134 comments.

For the sake of clarity, we list the abbreviations or acronyms of the participating institutions in Table 1.

Fig 1 shows the details of the participants by country of origin (left) and affiliation (right).

Fig 2 shows the number of participants who subscribed to (i.e., attended) the debate for a given paper and the number of interventions (questions, comments, etc.) in that particular debate. The data shown in this figure are surprisingly positive. It is rare to have more than 10 interventions in a face-to-face talk, but it was usual in this conference, with even more than 100 in some cases. This shows that in online conferences, asynchronous communication allows for deeper reflection when posing questions and also when answering them: Participants and authors can prepare their interventions more carefully.

Fig 3 shows the papers ordered by number of interventions. It is clearly observed that the minimum number of interventions made in 80% of the presentations was 19, while the minimum number of interventions made in approximately half of the presentations was 33. These figures are difficult to achieve in face-to-face conferences.

The attendee satisfaction data are shown in Fig 4. More than 90% of the attendees were satisfied or very satisfied with regard to access and ease of participation in the debates on the presentations and were willing to participate in future conferences of this type. In terms of satisfaction with the answers and their usefulness, the percentage was slightly higher than 80%, although this figure is still very positive.

The speaker satisfaction data are shown in Fig 5. Approximately 60% of the speakers were satisfied or very satisfied with the number of interventions and their usefulness; this percentage exceeds 60% in terms of the ease of responding to the issues raised. More than 50% think that online participation is more profitable than face-to-face participation, and almost 60% will

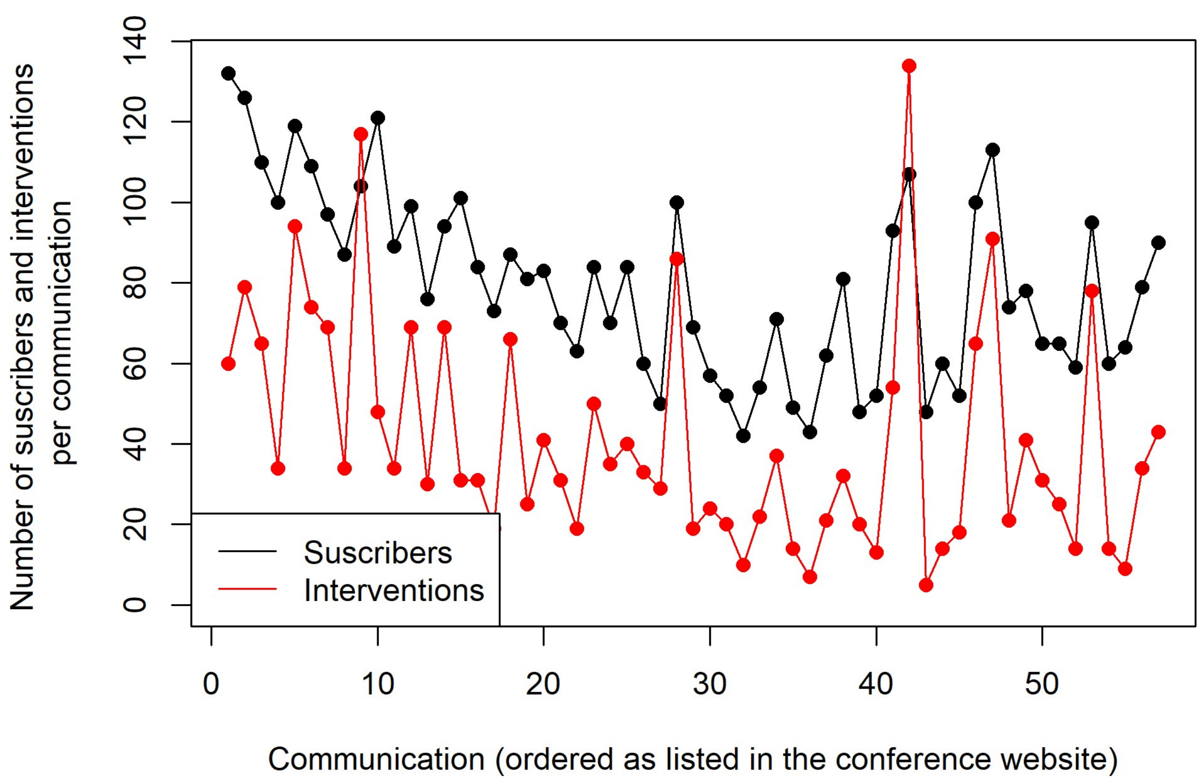



**Fig 2. Participants subscribed to the debates and number of interventions per paper.**

https://doi.org/10.1371/journal.pcbi.1007667.g002

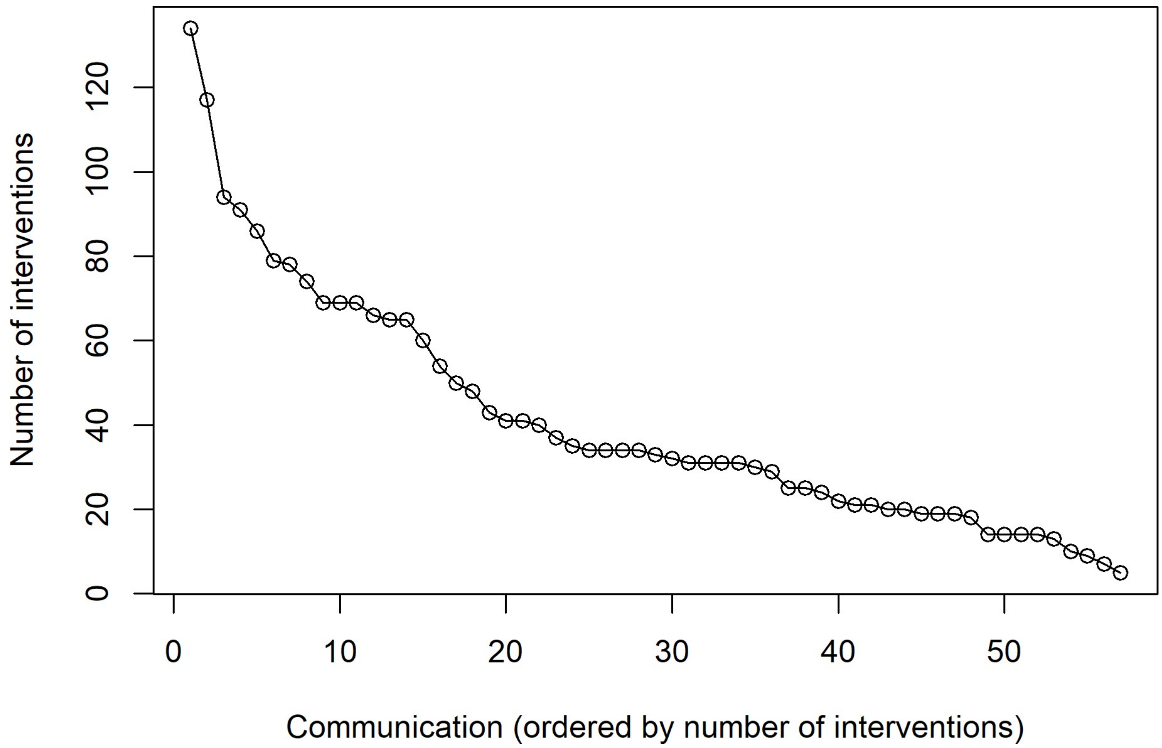



**Fig 3. Papers ordered by number of interventions.**

https://doi.org/10.1371/journal.pcbi.1007667.g003

probably present a paper in future online conferences. Even though these data are positive, they indicate that there is still room for improvement in the preparation and development of online conferences. One aspect that needs to be improved is to promote closer, more personal interactions in a synchronous manner.

Regarding the satisfaction of the scientific committee, it is worth mentioning that more than 70% stated (see Fig 6) that they had participated in virtual congresses before. This shows a substantial change in the trend of conference development and the enormous impact that it will have in the coming decade in terms of the use of ICT in the field of education. Nearly 100% of the respondents said they were satisfied or very satisfied with the review process. Regarding the quality of the papers, the committee considered that approximately 10% were very good quality and about 60% were good quality. Considering that one of the objectives of the conference was to be accessible to new researchers with few resources, the quality data can be considered satisfactory.

It is important to mention that online conferences are accessible to anyone who is interested in the subject. All that is needed to participate is an internet connection, so the participants' geographical location does not matter. In addition, accommodation and travel expenses are not required, and the registration fees are very low. They are even family friendly and respectful to the environment.

Our proceedings were published by the Communication and Publications Service of UJI. This is an editorial member of the Union of Spanish University Publishers (UNE), which guarantees the dissemination and marketing of this work nationally and internationally.

The publication, with ISBN: 978-84-17429-54-6, can be found in volume 19 of the Educational Innovation collection (see [22]) and can be downloaded freely in print version and to electronic devices.

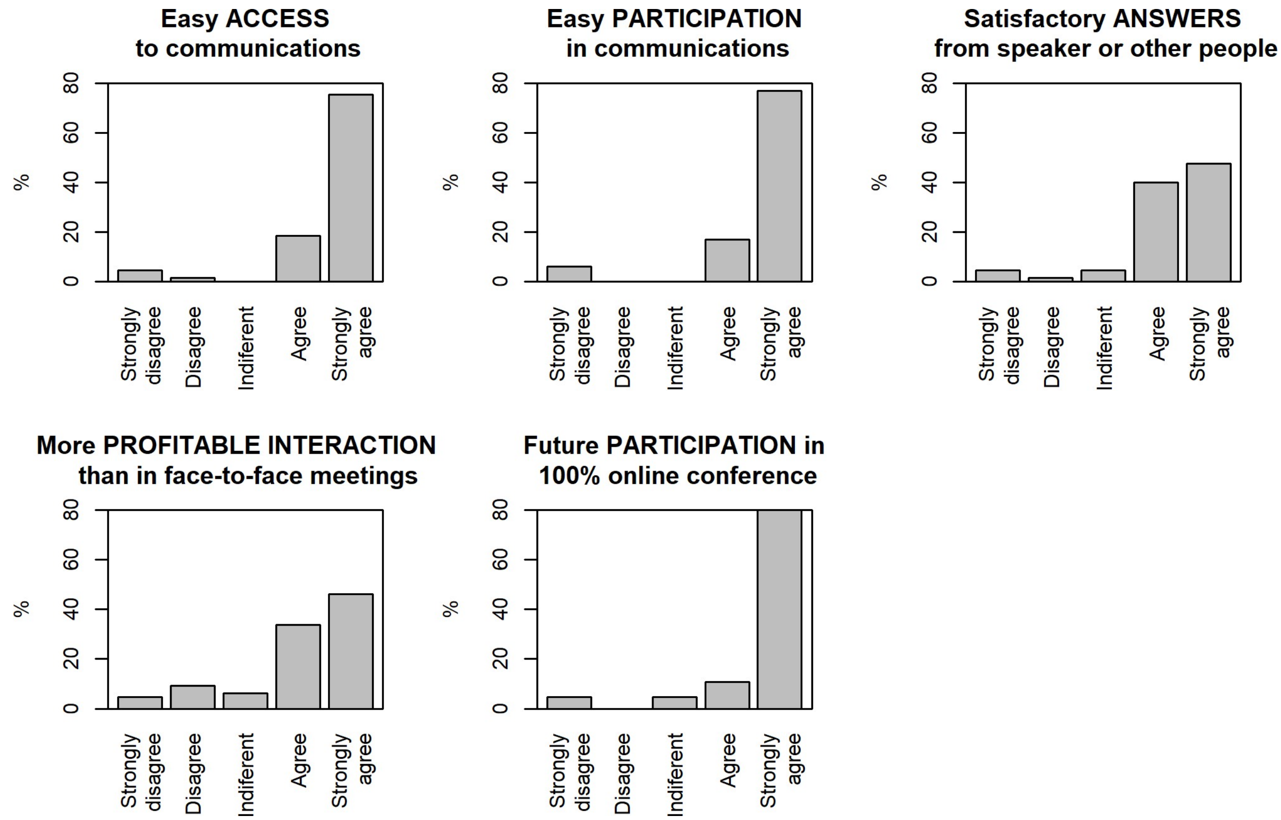



**Fig 4. Attendee satisfaction results.**

https://doi.org/10.1371/journal.pcbi.1007667.g004

Although we initially proposed 10 thematic areas to the authors, in this edition the works were finally grouped into four main thematic areas:

- Virtual environments: distance education, e-Learning, b-Learning, massive open online course (MOOC), etc.
- Skills assessment and planning: skills assessment, improvement of quality, planning of the European Credit Transfer System (ECTS), gender, legal and economic aspects of education, etc.
- Innovative experiences in education: methodologies, content, assessment, etc.
- New technologies in education: videos, apps, tablets, telephony, social networks, blogs, etc.

Readers who are interested in educational innovation and new ICTs applied to education can also check the proceedings from the first and second editions of the conference, held in 2016 (see [23]) and 2018 (see [22]), respectively.

In summary, in view of the quantitative and qualitative results, the experience has been very positive and rewarding. The satisfaction surveys also show that it was well received.

The success of a virtual conference depends on many factors. One of the most important factors is the planning: Committee meetings, creation of the virtual environment, and announcements on forums and social media. In addition, the choice of an online conference-management system, such as EasyChair, speeds up the workflow and was satisfactory for authors and committee members. The structure of the conference website with a different

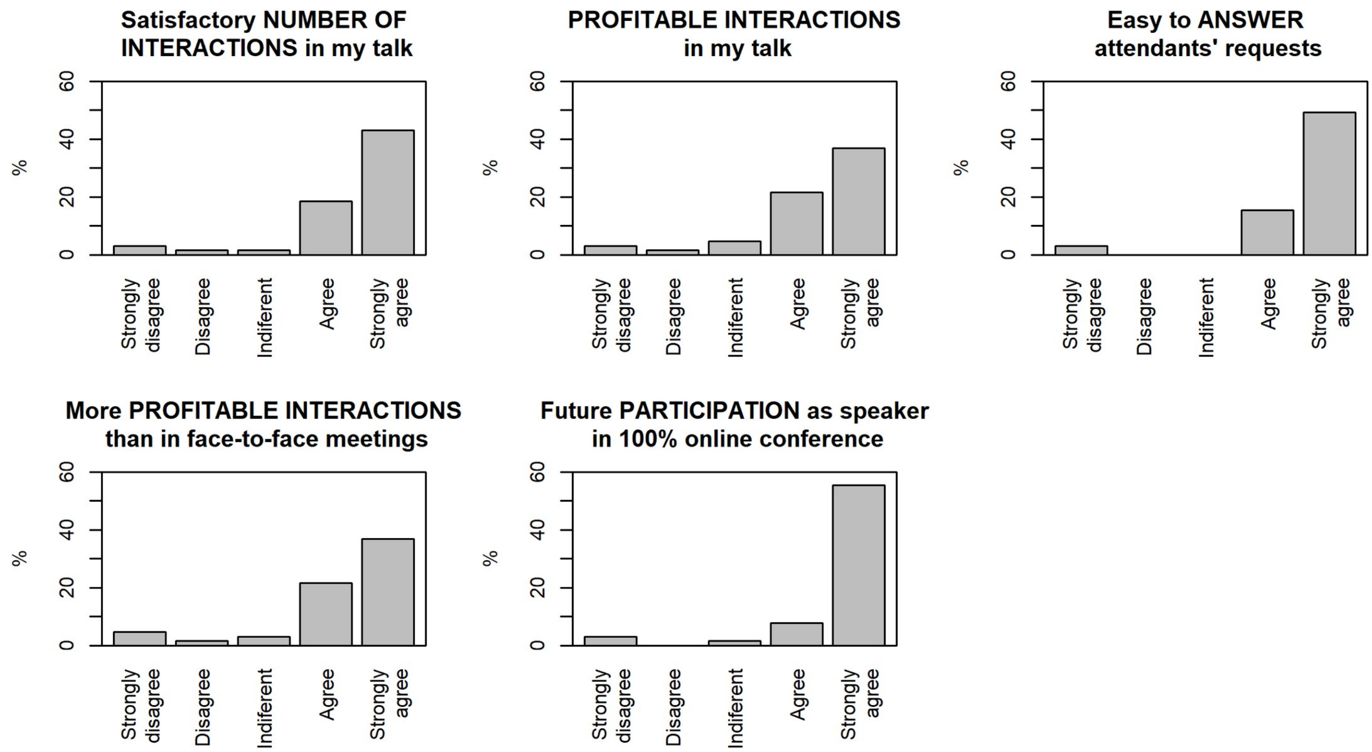



**Fig 5. Speaker satisfaction results.**

https://doi.org/10.1371/journal.pcbi.1007667.g005

forum for each talk provided quick and easy access to the discussions. Moreover, constant interaction between the participants increased their satisfaction.

Finally, as future work, we believe it is necessary to achieve better emulation of face-to-face contact in a synchronous way, as happens at face-to-face conferences. On this occasion, we

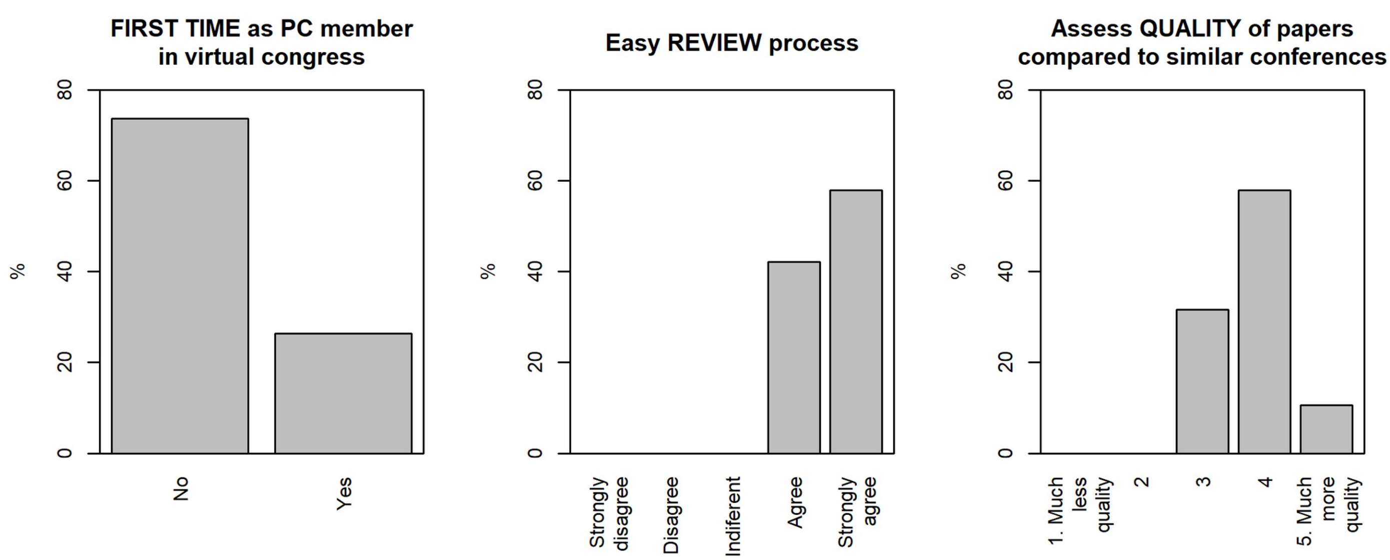



**Fig 6. Scientific committee satisfaction results.** PC, program committee.

https://doi.org/10.1371/journal.pcbi.1007667.g006

launched a direct YouTube connection in the form of a coffee lounge, but the reality is that participation was limited.

We think that enhancing this connection through constant encouragement by a host in charge of this task throughout the conference could be effective.

## References


1. Garrison DR, Cleveland-Innes M. Facilitating cognitive presence in online learning: Interaction is not enough. *The American Journal of Distance Education*. 2005; 19(3):133–148. https://doi.org/10.1207/s15389286ajde1903_2
2. Anderson L. *The theory and practice of online learning*. Athabasca University Press. 2nd ed. Canada; 2008.
3. Anderson L, Anderson T. *Online conferences*: *Professional development for a networked era*. Charlotte, NC: Information Age Publishing; 2010.
4. Bos AL, Sweet-Cushman J, Schneider MC. Family-friendly academic conferences: a missing link to fix the "leaky pipeline"?. Politics, Groups, and Identities. 2019; 7(3):748–758, https://doi.org/10.1080/21565503.2017.1403936.
5. Calisi RM, Working Group of Mothers in Science. Opinion: How to tackle the childcare–conference conundrum. *Proceedings of the National Academy of Science*. 2018; 115(12): 2845–2849. https://doi.org/10.1073/pnas.1803153115.
6. Anderson L, Anderson T. Online professional development conferences: An effective, economical and eco-friendly option. Canadian Journal of Learning and Technology. 2009; 35(2). http://dx.doi.org/10.21432/T29015.
7. Abdullah MH. Social Presence in Online Conferences: What Makes People 'Real'? *Malaysian Journal of Distance Education*. 2004; 6(2): 1–22.
8. Kear K, Chetwynd F, Jefferis H. Social presence in online learning communities: The role of personal profiles. Research in Learning Technology. 2014; 22. https://doi.org/10.3402/rlt.v22.19710.
9. Gichora NN, Fatumo SA, Ngara MV, Chelbat N, Ramdayal K, et al. Ten Simple Rules for Organizing a Virtual Conference—Anywhere. PLoS Comput Biol. 2010; 6(2): e1000650. https://doi.org/10.1371/journal.pcbi.1000650 PMID: 20195548
10. Corpas M, Gehlenborg N, Janga SC, Bourne PE. Ten Simple Rules for Organizing a Scientific Meeting. PLoS Comput Biol. 2008; 4(6): e1000080. https://doi.org/10.1371/journal.pcbi.1000080 PMID: 18584020
11. Budd A, Dinkel H, Corpas M, Fuller JC, Rubinat L, Devos DP, et al. Ten Simple Rules for Organizing an Unconference. PLoS Comput Biol. 2015; 11(1): e1003905. https://doi.org/10.1371/journal.pcbi.1003905 PMID: 25633715
12. Ware M. Online submission and peer-review systems. *Learned Publishing*. 2005; 18: 245–250. https://doi.org/10.1087/0953151054636471
13. EasyChair Conference System. http://www.easychair.org/. [cited 2019 Feb].
14. HotCRP conference management system. http://read.seas.harvard.edu/~kohler/hotcrp. [cited 2018 Feb].
15. Kanav S., Lammich P, Popescu A. A conference management system with verified document confidentiality. In Biere, A., & Bloem, R. Computer Aided Verification - 26th International Conference. Lecture Notes in Computer Science. 8559. Springer. 2014: 167–183. https://doi.org/10.1007/978-3-319-08867-9_11
16. ConfTool. https://www.conftool.net/ [cited Feb 2018].
17. OpenConf. https://www.openconf.com/ [cited 2019 Apr].
18. Lo SW, Phan RCW, Goi BM. On the security of a popular web submission and review software (WSaR) for cryptology conferences. In Kim S, Yung M, Lee HW. WISA 2007. Lecture Notes in Computer Science 4867. Springer. 2007: 245–265. https://doi.org/10.1007/978-3-540-77535-5_18
19. Lortie CJ. Ten simple rules for short and swift presentations. PLoS Comput Biol. 2017; 13(3): e1005373. https://doi.org/10.1371/journal.pcbi.1005373 PMID: 28358832
20. Pechakucha. https://www.pechakucha.com/cities/tokyo. [cited 2019 Apr].
21. Tatman R. Gender and dialect bias in YouTube's automatic captions. In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. 2017: 53–59.

22. Arnal AM, Barrachina S, Castelló J, Epifanio I, Galindo C, Gregori P, et al. Actas del congreso virtual: Avances en tecnologías, innovación y desafío de la educación superior ATIDES 2018. Castellón de la Plana: Publicacions de la Universitat Jaume I; 2018. http://dx.doi.org/10.6035/InnovacioEducativa.2018.19.

23. Arnal AM, Castelló J, Epifanio I, Galindo C, Gregori P, Lluch A, et al. Actas del congreso virtual: Avances en tecnologías, innovación y desafío de la educación superior ATIDES 2016. Castellón de la Plana: Publicacions de la Universitat Jaume I; 2016. http://dx.doi.org/10.6035/InnovacioEducativa.2016.16

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for organizing a webinar series

**Faisal M. Fadlelmola[1☯]*, Sumir Panji[2☯], Azza E. Ahmed[1,3☯], Amel Ghouila[4☯], Wisdom A. Akurugu[5☯¤a], Jean-Baka Domelevo Entfellner[6¤b], Oussema Souiai[7,8], Nicola Mulder[2], H3ABioNet Research working group as members of the H3Africa Consortium[¶]**

**1** Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Khartoum, Sudan, **2** Computational Biology Division, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa, **3** Department of Electrical and Electronic Engineering, Faculty of Engineering, University of Khartoum, Khartoum, Sudan, **4** Laboratory of Transmission, Control and Immunobiology of Infections (LTCII), Institut Pasteur de Tunis, Tunis-Belvédère, Tunisia, **5** Noguchi Memorial Institute for Medical Research, University of Ghana, Legon, Accra, Ghana, **6** South African MRC Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville 7535, Cape Town, South Africa, **7** Laboratory of BioInformatics Biomathematics and bioStatistics, Institut Pasteur de SalTunis, Tunis, Tunisia, **8** Institut supérieur des technologies médicales, Univesité Tunis al Manar, Tunis, Tunisia

☯ These authors contributed equally to this work.
¤a Current address: Computational Biology Division, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
¤b Current address: Biosciences eastern and central Africa (BecA-ILRI Hub), International Livestock Research Institute, Nairobi, Kenya
¶ Membership list of the H3ABioNet Research working group can be found in the Acknowledgments section.
* faisal.mohamed@hotmail.com



OPEN ACCESS

**Citation:** Fadlelmola FM, Panji S, Ahmed AE, Ghouila A, Akurugu WA, Domelevo Entfellner J-B, et al. (2019) Ten simple rules for organizing a webinar series. PLoS Comput Biol 15(4): e1006671. https://doi.org/10.1371/journal.pcbi.1006671

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** April 1, 2019



**Funding:** H3ABioNet is supported by the National Institutes of Health Common Fund grant number (U41HG006941). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

Francis Ouellette is a member of the H3ABioNet Scientific Advisory Board.

## Introduction

Technological advancements are rapidly changing the face of science in terms of data acquisition, its transfer, storage, analysis, interpretation and dissemination of results [1]. In biology and genomics, this is affecting many traditionally considered purely wet lab experiments like genome sequencing [2], medical diagnosis [3], and drug design [4]. It therefore becomes essential for bioinformaticians to remain up to date with recent trends and innovations in the field. Within Africa, this is even more true as the continent is striving to foster the development of innovative tools and strategies to improve health outcomes on the continent. H3ABioNet [5], the pan-African Bioinformatics Network, was established with the aim of capacity building in mind to further advance genomics research in Africa. It is therefore complementary to its other training initiatives [6] to ensure African scientists have access to avenues to disseminate their research, discuss their work, and network with peers.

Seminars and conferences are good opportunities for sharing and discussing new insights and networking with peers and can be considered as scientific meetings [7–13]. However, with prohibitive traveling costs and increased logistics, it is not always feasible to organize and attend numerous regular seminars. In an increasingly interconnected world brought about by technological advancements in communications, other alternatives can be used to supplement the in-person experience. Examples within Africa include: the H3ABioNet offering of a hybrid-delivery 3-months course, Introduction to Bioinformatics [18]. The Global Women in Data Science (WiDS) conference is another example of a one-day technical conference that is live-streamed from various locations across the globe (http://www.widsconference.org/). Mozilla Open Leaders (https://mozilla.github.io/leadership-training/), and global sprints (https://foundation.mozilla.org/en/opportunity/global-sprint/) are other examples of active engagement and community building that are arranged and conducted remotely.

Through organizing regular online seminars, known as webinars, H3ABioNet is aiming to empower a predominantly African audience to reap the benefits of being kept abreast of current research trends by expert domain scientists.

Webinars present a great virtual opportunity to engage and stimulate interactions between presenters and participants, and can accommodate more participants than a physical conference room setting which could be limited by space and accessibility [8]. Webinars provide participants the convenience of attending an academic presentation from the comfort of their offices or homes while multi-tasking. A successful webinar session is strongly dependent on the planning activities prior to the session.

We share in this paper ten simple rules for hosting a regular webinar series with particular emphasis on resource-constrained communities like many in Africa. These rules are derived from experiences gained and lessons learnt while organizing and running the H3ABioNet bioinformatics webinar series.

## Rule 1: Assemble an effective webinar coordination team

Similar to organizing any scientific meeting [7], hosting of a regular webinar series requires the involvement of a dedicated group of people. The role of the webinar coordination team is to assist with all the planning and logistics for hosting a webinar, while ensuring that the workload is not borne solely by a single individual. The involvement of postgraduate students and postdocs as part of the webinar coordination team facilitates the development of their skills in planning, communication, use of various conferencing platforms, coordinating and hosting of scientific events. These soft skills are vital for collaboration and working in large consortia which are not usually part of their normal training. To ensure the smooth running of the webinars, regular planning and post webinar meetings play an important role in iteratively developing and refining the planning procedures based on the challenges and successes of the previous webinars (Rule 10). This helps develop a cohesive and cooperative team of coordinators despite the fact that they might be located in different time zones and with busy schedules.

## Rule 2: Align a webinar theme to the expectations of the audience

Webinars are used as platforms to enable knowledge exchange, and to disseminate methods, results and best practices. They are generally aimed at specific audiences and address specific themes. Thus, choosing webinar themes requires mapping of the target audience needs and interests [7].

A webinar series could be instrumental in capacity building in various ways: 1) By hosting early career researchers, which gives them an opportunity to get early feedback on their research findings and practice scientific presentations skills. Given their limited experience, shorter durations (around 20 minutes talk each) are appropriate. 2) By hosting senior researchers in areas of interest to the intended audience. One senior scientist per session that lasts for 45-50 minutes are more reasonable. This enables the early career and junior researchers to expand their horizons and polish their ideas. It may also open avenues for collaboration or spring new research directions.

## Rule 3: Consider a webinar planning checklist

The main pre-webinar planning objectives and activities should be drafted and agreed upon by the webinar coordination team and a recurring list of tasks and responsibilities should be identified and drafted from the outset so each person's roles and responsibilities are clear

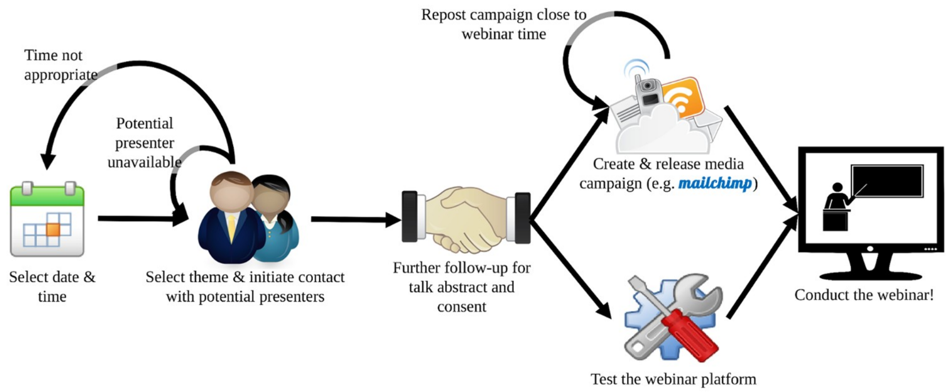



**Fig 1. Planning flowchart.** A webinar series flowchart of planning and logistics activities.

https://doi.org/10.1371/journal.pcbi.1006671.g001

(see Fig 1). It is important to include all the webinar team members in the identification and planning of the tasks from the start so they are more engaged as stakeholders and facilitate task distribution. The earlier tasks and timelines are identified, the better in terms of time available to prepare for the hosting of a webinar. Tasks include: determining the date for the webinar (rule 5), approaching potential speakers, obtaining the webinar talk, abstract and presenter's biography for creating the webinar announcement (rule 8), having a test run with the presenter (rule 9) using the chosen webinar platform (rule 6).

## Rule 4: Share webinar organizational documents

An accessible shared space for webinar documents enables decentralized access and smooth organization of tasks and resources. Various template documents on organizational letterheads can be created as the basis for gathering information from invited webinar speakers. Information collected should include the webinar title, authors, abstract, a picture of the speaker and their biography. An important document required is a consent release form whereby the speakers give their permission to use their submitted picture for the webinar announcement, record the webinar and acknowledge that they are the authors of the work. The webinar consent form also asks the speaker to choose which Creative Commons (CC) license they would like to link to the talk (CC by SA is the recommended license) and most importantly provides the speaker's permission to host the recorded webinar on various platforms such as a website or YouTube channel. The completion of these templates also allows the webinar coordinators to check with the speakers if there might be any sensitive unpublished results in the talk and discuss whether they should be presented. Various information captured through these templates are used to generate a webinar announcement (rule 8). For consistency, the same announcement format is used for all the webinars.

## Rule 5: Plan early and devise a calendar of regular activities

The regular cycle of hosting regularly occurring webinars inevitably translates into a series of recurring deadlines. Creating a calendar of webinar events helps mitigate the sudden onset of deadlines. In doing so, it is sufficient to settle on the provisional dates and times of the webinar events along with their themes. Speakers can be identified and confirmed later on accordingly (see Rule 7). It's important to pick the webinars time to accommodate audiences in various time zones. In today's increasingly interconnected world, there will inevitably be clashes with other meetings or workshops. Hence the earlier a regular webinar date and time is settled upon, the earlier it will make its way into attendees' calendars thereby enabling them to avoid subsequent scheduling conflicts. Another advantage of devising a calendar of activities earlier, is that due dates with reminders to perform specific tasks can be set up well in advance, themes settled upon and potential speakers contacted reasonably early enough for their availability to provide a talk before their calendars are filled.

## Rule 6: Settle on a convenient and user friendly webinar platform

The choice of platform for hosting a webinar is crucial. There are numerous free and paid-for platforms available offering different features. An important point to take into consideration within resource-limited settings is that most regions are operating on very constrained bandwidth and have limited budgets, so the use of expensive webinar platforms that have high bandwidth requirements and charge per user may not be feasible.

Free and commercial platforms exist that may be suitable for streaming webinars and other interactive Online events, such as: GoToMeeting, Zoom, Adobe Connect, Vidyo, Mconf and Google hangouts. These platforms vary in the features they provide, and it is of value to compare and assess before committing to a platform. Table 1 provides an example comparison of the key features offered in two platforms evaluated by the authors.

## Rule 7: Select theme expert presenters

The success of a good webinar hinges on the relevance of the talk to the proposed theme. As the various themes have been decided beforehand, the webinar coordination team members are able to reach out to other consortium members to identify speakers with relevant domain expertise. An advantage of having predetermined themes for a webinar is the flexibility to pre-identify more than one potential domain expert speaker.

In case of non-availability of the first potential speaker approached, other identified speakers can be approached. One of the webinar coordination team members drafts an invitation message to the potential presenter which includes the date and time of the webinar. Once an invited webinar speaker accepts the invitation to present, the template to capture the webinar abstract and the speaker's, biography and the webinar recording consent form are sent to the presenter with a due date for completion. A request is also made to provide at least three convenient dates and time slots to run through the logistics of the webinar platform being used and the format of the webinar.

Inclusion of all the coordination team members during communication with invited speakers will serve as a backup for any follow-up responses.

**Table 1. Comparison of the two webinar platforms used.**

| Comparison aspect | Google Hangouts | Mconf |
|---|---|---|
| Free/open source | Free | License |
| Ease of use | Could require Hangouts plugin installed; works best with Chrome browser | Requires Flash installed, works best with Mozilla Firefox browser. |
| Video & Screen sharing | Yes | Yes |
| Recorded videos | Readily available as Youtube videos | Available as html page files that requires rendering |
| Room size limit | 10 (free plan)—25 (business & education) | 50-60 |
| Privacy | Can join from Hangouts (with Gmail account), or from Youtube (Anonymously) | Login with anonymous guest name or registered mconf user |
| Notable problems | "Trying to reconnect" error | All traffic goes through central servers, that can be occasionally overloaded or down. |

https://doi.org/10.1371/journal.pcbi.1006671.t001

## Rule 8: Announce webinars through mailing lists and social media platform

Key to attracting a diverse and active audience is the dissemination of the webinar announcement earlier, and through both relevant mailing lists and social media channels that the target audience typically use (rule 2). The webinar announcements, containing all the required information: date, time, tool or web address, talk title, abstract and speaker's biography can be designed in marketing platforms like MailChimp for effective dissemination. Social sharing icons on the webinar announcement, use of relevant hashtags, and mentions of the speakers and their institutes all contribute to increased visibility of a given webinar announcement.

Twitter is one of the most used social channels for live webinars. You might consider running a live Q&A on Twitter in conjunction with your webinar. This can be done through designated hashtag or by posting questions to your audience and asking them to reply. It's good to see what your audience is saying to you and each other about your presentation. During the event, your moderator or social media coordinator should be highly visible on Twitter: answering questions, sharing interesting stats, engaging with attendees, and routing technical challenges.

## Rule 9: Allocate time for the platform orientation

Testing of the platform with the presenter is crucial to ensure a successful webinar session. This includes going through the webinar interface with the speaker to familiarise her/him by navigating through the presentation slides and testing different functionalities such as a mouse guided pointer.

The webinar flow is also discussed with the speaker during the test. Pre-testing also enables the webinar coordination team to assist the speaker fix any potential software or dependency issues before the actual webinar. Some platforms require initial add-ons or software setup which the presenter might not have installed. Depending on the platform used, sufficient details and instructions on how to test connection and/or software should be provided in the webinar announcement, usually in the form of web address to test the connection.

## Rule 10: Iteratively assess and evaluate what works and what doesn't

Running and maintaining scientific events like a recurring webinar series needs keeping careful track of metrics for regular assessment and evaluation [14]. Keep in mind that the online environment is still an evolving and untested media, and you will undoubtedly have to adopt a trial-and-error approach to find what works best [15].

In order to assess the webinars, a post webinar survey should be shared with all attendees. Such a survey could also ask for suggestions on themes or topics to be included in the webinar series.

Most of the webinar audience will normally go through the webinar advert and based on the advertised themes decide to attend a particular session. When such expectations are not met, webinars tend to have low participation. Webinars with pictures, graphs, tables and other diagrammatic representation get participants much more interested in following the session to the end than textual presentations [16, 17]. Webinars' presenters need to carefully develop a presentation that truly reflects the advertised themes while devising ways to sustain the audience throughout the session. More efforts in the training of early career researchers on quality and professional presentation is highly recommended.

## Summary

The webinars series form part of the regular activities of any scientific consortium that aims to strengthen research activities and foster collaborations amongst the different partners. The webinars are intended to foster the exchange of ideas, build potential collaborations across multiple disciplines and enabling the participation and sharing of knowledge in current research.

The Ten Simple Rules for organizing a webinar series can be summarized as follow: The webinar coordination team assists with all the planning and logistics for hosting a webinar (Rule 1); Choosing webinar themes requires mapping of the target audience needs and interest (Rule 2); Drafting a webinar planning checklist through regular planning meetings as well as post webinar meetings (Rule 3); Decentralized webinar organization of tasks and resources through accessible shared space (Rule 4); Planing early and settling on the provisional dates and times of the webinar events along with their themes (Rule 5); Choosing and settling on convenient and user friendly webinar platform (Rule 6); Approaching and confirming potential speakers (Rule 7); Obtaining the webinar title, abstract and presenter's biography for creating the webinar announcement through emails and social media channels (Rule 8); Allocating time for the platform orientation (Rule 9); and Keeping close-up track of webinar metrics for regular assessment and evaluation (Rule 10).

These Ten Simple Rules shared with the computational biology community will help those who have not yet ventured into training through webinars to learn from our experience. In our experience, the feedback from the post-webinar surveys clearly demonstrate that webinars are an effective way to create a two-way conversation between presenters and participants via a web-based platform.



## Acknowledgments

The authors would like to thank Suresh Maslamoney from CBIO at UCT and Kasandra Pillay from CSIR for Mconf technical support, along with the South African National Research Network (SANREN). The authors would also like to thank all the H3ABioNet webinar series' presenters for contributing toward the successful experiences and sharing their expertise skills with the consortium members.

H3ABioNet Research Working group members: Faisal M. Fadlelmola, Sumir Panji, Azza E. Ahmed, Amel Ghouila, Wisdom A. Akurugu, Jean-Baka Domelevo Entfellner, Oussema Souiai, Nicky Mulder, Sonia Abdelhak, Ezekiel Adebiyi, Imane Allali, Ahmed Mansour Alzohairy, Saaïd Amzazi, Bouabid Badaoui, Nicholas Bbosa, Alia Benkhala, Taoufik Bensellak, Ozlem Tastan Bishop, Gerrit Botha, Balkiss Bouhaouala-Zahar, Malek Chaouch, Bouchra Chaouni, Sara Eljadid, Segun Fatumo, Junaid Gamieldien, Simane Gaseitsiwe, Anita Ghansah, Atinuke Hassan, Scott Hazelhurst, Kerapetse Joromea, Ali Kishk, Almamy I. Koita, Daniel Masiga, Olfa Messaoud, Bruno Mmbando, Ahmed Moussa, Zahra Mungloo-Dilmohamud, Patrick Musicha, Trust Odia, Idowu Omotuyi, Hugh-George Patterton, Chaimae Samtal, Deogratius Ssemwanga, Gordon Wells.


## References


1. Bell G, Hey T, Szalay A. Beyond the data deluge. Science. 2009; 323(5919):1297–1298. https://doi.org/10.1126/science.1170411 PMID: 19265007
2. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. Annu Rev Genomics Hum Genet. 2016; 17:95–115. https://doi.org/10.1146/annurev-genom-083115-022413 PMID: 27362342
3. Avelino KYPS, Silva RR, da Silva AG Junior, Oliveira MDL, Andrade CAS. Smart applications of bionanosensors for BCR/ABL fusion gene detection in leukemia. Journal of King Saud University—Science. 2017. https://doi.org/10.1016/j.jksus.2017.08.002

4. Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther. 2013; 138 (3):333–408. https://doi.org/10.1016/j.pharmthera.2013.01.016 PMID: 23384594
5. Mulder NJ, Adebiyi E, Alami R, Benkahla A, Brandful J, Doumbia S, et al. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. Genome Res. 2016; 26 (2):271–277. https://doi.org/10.1101/gr.196295.115 PMID: 26627985
6. Aron S, Gurwitz K, Panji S, Mulder N, Consortium H. H3ABioNet: developing sustainable bioinformatics capacity in Africa. EMBnet j. 2017; 23(0):886. https://doi.org/10.14806/ej.23.0.886
7. Corpas M, Gehlenborg N, Janga SC, Bourne PE. Ten simple rules for organizing a scientific meeting. PLoS computational biology. 2008; 4(6):e1000080. https://doi.org/10.1371/journal.pcbi.1000080 PMID: 18584020
8. Manallack DT, Yuriev E. Ten simple rules for developing a MOOC. PLoS computational biology. 2016; 12(10):e1005061. https://doi.org/10.1371/journal.pcbi.1005061 PMID: 27764089
9. McInerny GJ. Ten Simple Rules for Curating and Facilitating Small Workshops. PLoS computational biology. 2016; 12(7):e1004745. https://doi.org/10.1371/journal.pcbi.1004745 PMID: 27441642
10. Pavelin K, Pundir S, Cham JA. Ten simple rules for running interactive workshops. PLoS computational biology. 2014; 10(2):e1003485. https://doi.org/10.1371/journal.pcbi.1003485 PMID: 24586135
11. Searls DB. Ten simple rules for online learning. PLoS computational biology. 2012; 8(9):e1002631. https://doi.org/10.1371/journal.pcbi.1002631 PMID: 23028268
12. Budd A, Dinkel H, Corpas M, Fuller JC, Rubinat L, Devos DP, et al. Ten simple rules for organizing an unconference. PLoS computational biology. 2015; 11(1):e1003905. https://doi.org/10.1371/journal.pcbi.1003905 PMID: 25633715
13. Gichora NN, Fatumo SA, Ngara MV, Chelbat N, Ramdayal K, Opap KB, et al. Ten simple rules for organizing a virtual conference—anywhere. PLoS computational biology. 2010; 6(2):e1000650. https://doi.org/10.1371/journal.pcbi.1000650 PMID: 20195548
14. Bourne PE, Barbour V. Ten simple rules for building and maintaining a scientific reputation. PLoS computational biology. 2011; 7(6):e1002108. https://doi.org/10.1371/journal.pcbi.1002108 PMID: 21738465
15. Bik HM, Dove AD, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. Ten simple rules for effective online outreach. PLoS computational biology. 2015; 11(4):e1003906. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439
16. Bourne PE. Ten simple rules for making good oral presentations. PLoS computational biology. 2007; 3 (4):e77. https://doi.org/10.1371/journal.pcbi.0030077 PMID: 17500596
17. Lortie CJ. Ten simple rules for short and swift presentations. PLoS computational biology. 2017; 13(3): e1005373. https://doi.org/10.1371/journal.pcbi.1005373 PMID: 28358832
18. Gurwitz Kim T., Aron Shaun, Panji Sumir, et al. Designing a course model for distance-based online bioinformatics training in Africa: The H3ABioNet experience. *PLOS Computational Biology*, 13(10): e1005715, 2017. https://doi.org/10.1371/journal.pcbi.1005715 PMID: 28981516

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for measuring the impact of workshops

**Shoaib Sufi[1☯]*, Aleksandra Nenadic[1‡], Raniere Silva[1], Beth Duckles[2☯], Iveta Simera[3‡], Jennifer A. de Beyer[3], Caroline Struthers[3], Terhi Nurmikko-Fuller[4‡], Louisa Bellis[5‡], Wadud Miah[6‡], Adriana Wilde[7‡], Iain Emsley[8], Olivier Philippe[9], Melissa Balzano[10], Sara Coelho[11], Heather Ford[12], Catherine Jones[13‡], Vanessa Higgins[14‡]**

**1** School of Computer Science, University of Manchester, Manchester, United Kingdom, **2** Portland State University, Portland, Oregon, United States of America, **3** UK EQUATOR Centre, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom, **4** College of Arts and Social Sciences, Australian National University, Canberra, Australia, **5** Cancer Research UK Cambridge Centre, University of Cambridge, Cambridge, United Kingdom, **6** Numerical Algorithms Group, Oxford, United Kingdom, **7** School of Computer Science, University of St. Andrews, St. Andrews, Scotland, United Kingdom, **8** Oxford e-Research Centre, University of Oxford, Oxford, United Kingdom, **9** Electronics and Computer Science, University of Southampton, Southampton, United Kingdom, **10** ELIXIR Hub, Wellcome Genome Campus, Cambridge, United Kingdom, **11** EGI Foundation, Amsterdam, the Netherlands, **12** School of Media and Communication, University of Leeds, Leeds, United Kingdom, **13** Scientific Computing Department, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot, United Kingdom, **14** School of Social Sciences, University of Manchester, Manchester, United Kingdom

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* shoaib.sufi@manchester.ac.uk

OPEN ACCESS

**Citation:** Sufi S, Nenadic A, Silva R, Duckles B, Simera I, de Beyer JA, et al. (2018) Ten simple rules for measuring the impact of workshops. PLoS Comput Biol 14(8): e1006191. https://doi.org/10.1371/journal.pcbi.1006191

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** August 30, 2018



**Funding:** SS, AN, RS, IE, and OP acknowledge the support of EPSRC, BBSRC and ESRC Grant EP/N006410/1 for the UK Software Sustainability Institute, http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/N006410/1. IS, CS, and JdB acknowledges support from Cancer Research UK (grant C5529/A16895). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Workshops are used to explore a specific topic, to transfer knowledge, to solve identified problems, or to create something new. In funded research projects and other research endeavours, workshops are the mechanism used to gather the wider project, community, or interested people together around a particular topic. However, natural questions arise: how do we measure the impact of these workshops? Do we know whether they are meeting the goals and objectives we set for them? What indicators should we use? In response to these questions, this paper will outline rules that will improve the measurement of the impact of workshops.

## Introduction

The idea for this paper came from a workshop entitled 'Measuring the Impact of Workshops' [1]. 'Measuring the Impact of Workshops' collected practices and ways of thinking from a diverse set of experienced workshop organisers. This paper summarises these ideas into a coherent set of recommendations, 10 simple rules, that should make measuring impact more straightforward and more intentional.

Why should we measure the impact of the workshops we organise and run? With good measurements, we can convince funders to maintain and support the work that we do, encourage people to attend, and feel satisfied that the work that we are doing with our workshops is worthwhile and making a positive difference.

A consistent approach to measuring similar workshops allows us to compare them over time and show improvement or the need for adapting the workshop to be more successful for the intended audience.

Effective measurement is the precursor to evaluating the workshops that we organise, which allows us to make quality assertions, i.e., that our workshops deliver the benefits to stakeholders (funders, attendees, and ourselves) that we think they do. Workshops to provide training or information dissemination are a recognised communication and engagement activity that funders classify as pathways to impact [2] [3]. Impact is becoming increasingly important for assessing research; for example, the United Kingdom Research Excellence Framework (REF) has increased its weighting for impact from research from 20% to 25% for REF 2021 [4].

This paper focuses on measuring impact and the thinking, knowledge, skills, and techniques surrounding this. There are other excellent resources for organising [5] [6], curating, facilitating [7], and improving interactivity [8] at workshops, meetings, and conferences [9]. The reader is encouraged to consult them for broader information about effective workshop organisation and running.

This paper proposes 10 simple rules for measuring the impact of workshops. Rules 1 and 2 concern planning: what you need to think about to set the right context for being able to measure impact. Rules 3–5 are knowledge- and skill-based rules: things to be aware of or know how to do before constructing the method of measuring impact. Rules 6–10 are techniques that can be used to improve how we measure workshop impact.

## Types of workshops

This guide is focused on three types of workshops, which are explained below for reference and to provide consistent terminology. Rules 1–6, 8, and 10 apply to all three workshop types. Rules 7 (to understand a change in skills) and 9 (to assess skills learnt) are of particular interest and use for learning workshops, although they could also be used for the other workshop types. We will also illustrate, when necessary, which rules are of particular use for specific workshop types.

### Exploratory workshops

In exploratory workshops, ideas are analysed to better understand a topic and its associated problems, current solutions, and future challenges. These workshops can have aims such as identifying what actions are needed to move a particular topic forward or getting expert advice from and into different communities. The keynotes, lightning talks, miniworkshops, and discussion sessions at the Collaborations Workshop [10] series are an example of exploratory workshop sessions.

### Learning workshops

In learning workshops, a particular skill set, application, or technique is taught. The expected outcome is increased knowledge, competence, or confidence in a particular area or set of techniques. Examples of learning workshops are the Software Carpentry [11] and Data Carpentry [12] workshops. Such workshops typically include practical exercises to apply the knowledge gained with assistance provided by the workshop organisers.

### Creating workshops

Creating workshops bring together individuals with a common or intersecting interest to solve particular problems by collectively building something. They can include multidisciplinary teams in which problem holders guide the creative process. What is made can vary; it could be software, standards, resources, or even papers. Workshops in the humanities, in which collections of researchers work on the translation or annotation of historical texts, are more akin to creating workshops than traditional exploratory or learning workshops. The commonly termed 'hackathons' [13] are considered creating workshops for the purpose of this paper.

## Rule 1: Setting goals effectively

When developing your workshop, the workshop goals or objectives [14] need to be set. Define the outputs (what you want to produce) and outcomes (what difference you hope it will make) [15] for the workshop.

Although goal-setting is, in general, good workshop organisational practice, we highlight it here, as it is the foundation on which impact can be measured as we assess to what extent we met our stated goals.

One way of effectively measuring impact is to include those who will attend the workshop when creating the ultimate goals. By using the answers from a preworkshop questionnaire, the attendees can influence the creation of the goals. Their answers can help set goals such as the change in skills needed, topics that should be explored, and problems that need addressing. This can then guide what questions will be asked postworkshop in order to measure the overall impact.

Collecting feedback during workshop development could lead to small but significant adjustments in the programme to help meet workshop goals.

## Rule 2: Balancing time, effort, and costs

It is important to consider the cost of a workshop in terms of time, effort, and money when thinking about measuring its impact. The number of people, the duration, the venue (whether held in person or online), the price to the individual, the resources available within the organisation, and whether the workshop is one of a series or a stand-alone event can all affect how much effort is reasonable to put into measuring impact and which of the rules below are applied. For example, the impact of a free one-hour online workshop might be adequately analysed using a few survey-type questions sent out at the end of the workshop. However, a multiday, moderately expensive workshop that uses a mixture of learning and exploring and intends to enthuse people about changing practices may require more effort to be put into the impact analysis, so more of the rules would come into play (especially Rules 6–10, related to techniques).

It is expensive to fully measure long-term impact, such as how people are applying what they have learnt or to establish a causal link between workshop attendance and improved research (see Behaviour and Results in the Kirkpatrick model [16]) [17]. However, it is better to do at least some work in this space, imperfect as it may be (Rule 8), rather than insisting on measuring things perfectly or not at all [18].

## Rule 3: Create metrics purposefully

The process of taking a concept and converting it into a metric is called commensuration. Any time that we quantify something that is not easily turned into a metric, such as an idea like 'satisfaction' or 'comfort', we are engaging in commensuration [19]. Examples of commensuration include creating workshop evaluations, measures of job productivity such as human resource documents, and cost/benefit analyses.

Metrics help to make things more comparable, simplify complex information, and create standards that support easier decision-making. However, we should be aware of the context and assumptions made when a specific metric is created.

The metrics we form are ultimately made up of what we believe is important; we are a part of what we are trying to study. Therefore, metric formation (i.e., commensuration) has inherent bias. We often find ourselves measuring that which is easy to measure or that we most want metrics on. Being aware of this limitation allows us to be more honest and intentional about trying to minimise bias. By knowing what we value and what is easy to measure, we can examine our analysis and check where we are missing data.

When measuring the outcome of workshops, it is important to ask questions that will elicit useful responses to help us answer our research questions or goals and support our intended analysis of the results of the question. This does not mean that we should bias the research towards a particular end. Instead, we should gather data that is useful for the task of discovering the concepts and outcomes that matter most to us, whether through scoring, categorisation, or free-text responses.

## Rule 4: Understand bias

The work of controlling biases is never finished. An iconic study in the field of management sciences, the Hawthorne effect [20], showed that the act of studying other humans will affect the outcome of the study. We can only evaluate our data honestly if we know what our biases are and are willing to be open about where they might be coming from.

### Common biases

Table 1gives a nonexhaustive list of the common biases that can affect impact measurement [21].

### Controlling for bias

To control for bias, consider which biases will most likely affect the results of your study and determine strategies to counteract those biases to the best of your ability. Be conscious of the fact that bias always exists and consider how it will affect your analysis. For example, use best practices in asking questions in survey research (Rule 5).

**Table 1. Common biases and countermeasures.**

| Bias type | Explanation | How to counter |
|---|---|---|
| Confirmation bias | The tendency to reaffirm your own values and beliefs and to create research methods that confirm what you already believe to be true. For instance, I might decide that I'd like evidence that my workshops are very effective, so I ask questions designed to get mostly positive responses. | Know what we believe to be true and make certain that the questions allow for the opposite (and other) responses. |
| Sampling bias | When the sample you are drawing from is not representative of a larger population. Unless you get responses from every single person in a workshop, for instance, you will have a biased sample. For example, I might send out a workshop evaluation survey on a day when a third of the workshop attendees are at a conference so are not able to respond. | Check whether responders had similar profile distributions to those who attended the workshop. Compare demographics (gender, domain, career stage, etc.) to help detect bias even in anonymous surveys. However, such information could be used to identify individuals in a smaller workshop. |
| Social desirability bias | A person responding to questions wishes to give a response that will make the interviewer think well of them. For example, I might feel uncomfortable answering the question 'After this workshop on measuring impact, I feel confident about measuring the impact of my next event', if after the workshop I still didn't understand the topic. | Questions can emphasise the need for honesty and promise that although answers will be used and published, respondents will remain anonymous. For questions that ask about skill levels before and after a workshop (e.g., Rule 8), it is very important to indicate that it is okay if the respondent does not know how to do a skill. |

https://doi.org/10.1371/journal.pcbi.1006191.t001

## Rule 5: Design your surveys well

As part of a wider range of social research methods [22], surveys are a key mechanism for evaluating workshops. They can form part of the information gathering before people attend (e.g., during registration), at the workshop (e.g., for conferences [23]), after the workshop (e.g., as feedback forms), and much later after the workshop in follow up or impact surveys (see Rule 8).

### A note on quantitative versus qualitative

Two types of survey questions can be asked: quantitative and qualitative. Quantitative questions are usually answered by many respondents and have definitive answers. They often use Likert scales, in which respondents indicate how much they agree or disagree with a statement by choosing from a set of fixed choices on a linear scale (e.g., strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree) [24]. Qualitative questions are more open ended, and the answers can be probed using thematic analysis [25]. Answers to qualitative questions allow you to gather information about the workshop and formulate hypotheses. They can even guide you as to which quantitative questions you could ask in the future by helping you to identify the concepts your current questions are missing.

### Common pitfalls and how to avoid them

The questions asked and how they are constructed are an important part of survey design. Table 2 details what to watch out for when constructing survey questions to decrease bias and

**Table 2. How to decrease bias and increase clarity in survey questions.**

| Aspect | Explanation of issues | How to counter |
|---|---|---|
| **Compound question** | They are complex, overly wordy, and have multiple potential answers. An example of such a question is the following: '*Would you prefer if this workshop were offered on Thursdays for two hours or Fridays for two hours, or do you not care which day it's offered but only that it's offered every week*?' These questions are hard to answer clearly, and often only one portion of the question will be answered. Analysing the answers is thus potentially meaningless. | Deconstruct the compound question into separate questions. |
| **Leading question** | These questions guide the respondent toward a particular desired response. In combination with the social desirability bias, this is one of the easiest ways for survey research to become biased. For example, '*Given the number of people who have expressed an interest in weekend workshops, how interested are you in signing up for the workshop*?' | Remove any leading parts to the question. '*How interested are you in signing up for the workshop*?' |
| **Complex question** | These questions are a challenge for the respondent to follow and accurately respond to. Similar to the compound question, it makes it hard both for the respondent to answer accurately and for the researcher to know what is being measured. For instance, '*Imagine you are trying to teach a student who has never used the command line to do a pull request in GitHub. What are the ways you would teach that to someone from a different background than your own and how would you relate that to the teaching you would do of a loop in Python*?' | Pretest your survey so that these types of questions can be highlighted and reworded before you run the survey for real. |

(*Continued*)

**Table 2.** (Continued)

| Aspect | Explanation of issues | How to counter |
|---|---|---|
| **Multiple-choice question** | Multiple-choice questions that do not offer all of the possible answers included are naturally difficult to accurately respond to. For example, a question asking for a report of eye colour that does not include the respondent's eye colour in the possible answer choices cannot be answered. | Undertake qualitative work and/or pretest your survey to find all of the possible answers to your multiple-choice questions.<br>Include an 'other' response text box to capture other categories. Some manual coding and/or cleaning [26] will be needed to make use of the data. |
| | The order of the multiple-choice answers should be intuitive and have a flow. In some cases, it might make sense to randomise the choices to control for bias. In other situations, in which confusion could be caused (e.g., standard lists of ethnicity or domains), keeping a logical order is less confusing. Confusing those who fill in the survey is a sure way of decreasing response rates. | Check whether the answer choices should be randomised or kept in a logical/standard order. |
| **Wording choice** | It is rare that the use of absolutes such as 'always' or 'never' will help you write an effective survey question. Using an absolute in a survey question can mean that the response is not as useful because the respondent may have one instance that rules out an answer, e.g., '*Do you prefer workshops to always be run on Tuesdays*?' | In the majority of cases, remove or replace any absolute word(s) in questions. |
| | Keep answers comparable between respondents. For example, asking a respondent if they travelled 'far' to attend the workshop could be subjective, with some people considering 10 miles to be far and others considering over 100 miles to be far.<br>It is equally important to manage value-laden words, such as 'good' and 'bad'. However, value-laden and subjective questions can be useful for qualitative analysis of the workshop and can help to understand respondents' perspectives, leading to future quantitative questions. | Define what you mean when asking about matters that are open to subjective opinion, e.g., rather than 'far', you could give a selection of distances. 'Good' could be replaced with something more specific about your intent, such as 'useful' or 'enjoyable', depending on what you are trying to measure. |
| **Open-ended question** | Not offering one open-ended question can cause you to lose out on information from attendees. | The question allows respondents to highlight anything positive or negative about the workshop that they would like to bring up. This can act as an additional safety net to catch issues with the survey that may have slipped through pretesting. |

https://doi.org/10.1371/journal.pcbi.1006191.t002

increase the clarity of what is being asked, thereby improving the quality of the results. The overall goal is to design straightforward questions that respondents can easily understand and answer. A common response by participants to any barrier to answering a question, from technical difficulties or confusingly worded questions, is to not complete the survey.

## Rule 6: Ask about participants' 'confidence'

A common question that you can ask both at the start and end of a workshop is 'How confident are you about [workshop topic]?' This question allows you to gauge the participants' change in confidence and analyse whether the workshop changed the level of confidence about a particular subject, technique, or way of working together (e.g., creating workshops).

It is possible that participants' confidence might actually decrease as they realise that they know less than the other participants or they discover that there is much more to a particular field than they first realised. However, we only rarely found this to be the case in the Collaborations Workshop [10] series run by the Software Sustainability Institute [27]. On average, confidence levels at registration compared to confidence levels in the postcourse feedback showed an increase. If your participants report decreased confidence, look closer at the reasons why by mining the responses to open-ended questions or by following up with the participants. You could include an open-ended question asking about the direction and cause of the change in confidence if you suspect increased understanding may decrease confidence.

Asking about confidence has some limitations. For example, those answering the precourse question about confidence level may not be the same people answering the postcourse question, which prevents you from getting a true representation of the change in confidence level. You can mitigate this issue by asking participants to gauge their confidence level both pre- and postcourse in the same postcourse survey. You can then compare the average change across the same set of participants.

The term confidence can mean different things to different people; it might mean "how well you know the area" or "how well you can do something" or "how well you can explain something." This is okay, as what we are looking at is an individual's perceived change in confidence, whatever that means for them. If you are interested in just one of these possible interpretations, rephrase the question or add another question to ask about competence as well as confidence so that you can capture changes related to overall skills in an area, for example.

Asking about confidence is helpful if you want to know whether your workshop has made a difference for a particular field, area, or technique. If your goals require you to measure the change in your participants' skills, then Rule 7 (ask about specific skills) and Rule 9 (test specific skills) will be more important for you. These two rules are especially relevant measures for learning workshops.

## Rule 7: Ask about specific skills

As explained in Rule 1, all workshops should have objectives set. Objectives for learning workshops tend to be acquiring or mastering of specific skills or techniques. Objectives for exploring workshops tend to be knowledge, understanding, and an idea of where to look for more information or to find collaborators. Objectives for creating workshops tend to be learning a new skill, feeling like you have contributed toward a project, being able to do things differently, or finding future people to work with.

Although it is easy to ask questions around improvement in confidence (Rule 6), these questions are often too broad on their own. For deeper insights into the workshop's impact on its attendees, we need to craft more in-depth questions geared specifically to measuring the objectives (learning or otherwise) for those attending. You could ask people about their different levels of agreement for specific skills after the workshop using a Likert scale (e.g., strongly agree, agree, neutral, disagree, or strongly disagree).

Examples of specific skill questions are the following:

- I understand the purpose of [a particular technique].
- I can describe the [process].
- I can apply the [technique] to my work.
- I have a firm plan for how I am going to introduce what I have learned from this workshop into my work.

It can be difficult to repeat these sorts of questions in surveys immediately after the initial postworkshop survey and in a six-month postworkshop follow-up survey, as they are specific enough for participants to have forgotten the details. However, using the 'write to their future self' approach mentioned in Rule 8 can help to remind participants of their planned objectives before sending them a follow-up survey.

## Rule 8: Gather feedback before, during, and after

A postcourse feedback questionnaire is not the only way to measure the impact of a workshop. There are a number of other times in the process when asking participants questions can help to both run a more effective workshop and measure impact.

### Before

When running a workshop, it is important to collect demographic information of registrants, such as domain and career stage, to ensure that your audience is representative of the people you are targeting. This can be done at the precourse stage. Other specific data to capture are their learning expectations, what they hope to discuss during the workshop, and what their (perceived) existing competencies are in the subject. This information is useful for structuring the workshop and adapting the content to align with the registrants' needs.

A good time to gather this precourse information is at the point of registration. Participants are keen to attend the event and have a clear idea of why they are signing up and are thus likely to provide what is asked.

### During

During multiday events, feedback of how the event is going can be collected at the end of each day and fed back to facilitators and organisers. This kind of ongoing feedback allows you to identify and respond to problems as they occur. You can also keep a running score of how well participants feel the event is meeting its objectives or even see whether participants change their goals or what they feel the objectives of the event should be as the event goes on.

### At the end

At the end of a workshop, participants are normally bubbling with ideas, techniques they have learnt, things they want to change about their work upon their return, and which of the people they have met they might follow up with. However, normal life can sometimes take over, with pressing deadlines and the same old environment distracting the participants from carrying out their plans.

An excellent exercise is to ask participants to write to their future selves [28]. In this technique, the positive change envisaged by those who attended the workshop is captured in written form at the moment they are most enthusiastic. An example of how to run this exercise is to ask the participants during the last session of the workshop to write a postcard to themselves as a reminder for future actions. You could ask the participants to write about how they want to use what they have learned, how they would like to change some aspects of their current practice as a consequence of attending the workshop, or what their action plan is.

The postcards are then collected up by the organiser. Sometime after the workshop has closed, say two months to four months later, these postcards can be sent out to each of the participants as a reminder of what they planned. There is something intriguing about physical postcards in the age of digital communication, which only adds to the impact of such practices.

While evaluation questionnaires can help measure the impact of workshops, this technique is a fun but innovative way to extend the impact of the workshop beyond the time in which it was run.

### After

Postcourse feedback should be collected soon after the workshop is completed, while things are still fresh in people's minds. Ideally, if the survey is hosted online, the participants can be given access to the link during the course, reminded in person on the last day, and then reminded again by e-mail within two days, thereby maximising the chance of responses.

### Much after

Assessing the long-term impact and influence of your workshop on the individual can be difficult. To assess if your workshop has made or is making a difference, send out a survey sometime after your workshop has completed (e.g., four to six months after the workshop). You can ask questions about what the participants learned at the workshop, how they have applied this knowledge to their work, and what impact the knowledge and network has had on their working life and practices. Another option is to conduct one-on-one interviews with participants. Although time consuming to conduct, those who are willing to talk can offer a lot of useful information and a much more nuanced view of impact than a survey.

Some workshops will involve a cohort of participants that remain linked together after your workshop. For instance, your workshop might form one of a series that the participants will all attend, or they might attend your workshop as part of their degree training as a group. The participants might even form themselves into a cohort that did not exist before the workshop, choosing to remain connected after the workshop through regular meet-ups. Such cohorts can make it easier to get feedback. For cohorts that are formed before your workshop, you can factor in when they will meet to plan the intervals at which you will get feedback. For example, you could arrange recorded interviews with a selected number of participants from the cohort during one of their scheduled meetings as another way of collecting feedback. The recordings can then be used to promote the workshop, maximise its impact, and provide evidence to funders of how people used what they learned.

## Rule 9: Harness gamification to test participants' skills

In Rule 7, we asked participants whether they felt that they had acquired certain skills. We can also test whether they have acquired these skills. One way to assess if people have learned a particular skill from your workshop is to assess them indirectly through an informal learning and assessment platform. Asking the participants to play a game alleviates the features of standardised-testing environments that can cause anxiety. In games, learners encounter materials in new ways and have to apply their learning, not just repeat memorised details, and must rely on tacit knowledge. Games can show whether they have understood core concepts and knowledge areas. They can also highlight gaps and thus better focus the efforts of future workshops. For all its benefits, game playing remains underused, although some examples can be pointed to as useful case studies.

Such complementary assessments fall into the category of 'serious games' or 'games with a purpose' [29], with an example being the Treasure Explorers [30]. This tool combines different question types (multiple choice, tagging, and connecting ideas) as a way to help quantify people's understanding. The system was evaluated using games created to test understanding of logic and language, following the Brazilian National Educational Plan. People who use the system don't feel like they are being formally tested. The system also has a social element, which shows a leader board (connected to players' Facebook accounts) comparing how well players have done and allowing players to post their scores on social media. This competitive element again adds to what is termed 'playful learning'.

Developing a game-playing assessment system from scratch can be time consuming, but there are toolkits that can help [31] [32]. For a longer running workshop or a training series, it could form a worthwhile part of the evaluation method. Given the nature of such a system, it could even work with Rule 8; players (learners) can be sent a reminder to play, perhaps on a monthly basis, to keep the knowledge fresh in their minds and encourage them to use it in their day-to-day work [33]. A useful output from a creating workshop might be to make such systems for education, assessment, or even to solve parts of computational pipelines in their domains [34].

## Rule 10: Measuring those who did not attend

It is easy to forget to measure the impact your workshop had on those who did not attend. In today's social world, both organisers and participants have even more ways of sharing their content and what they have learnt 'beyond the room.' Live recordings, blog posts, tweets, Instagram photographs, and stand-alone reports are all ways to allow your workshop to keep reaching new audiences beyond the workshop date. Encourage your attendees to share their experiences during and after your sessions. Twitter is currently a favoured platform for such event amplification [35]. If you want to encourage event amplification, use a uniform hashtag across promotional material and, resources permitting, have one of your organisers actively contribute to and monitor the conversation during the workshop. This will increase the workshop's impact and your interaction with those who are not in physical attendance. Take the time to produce a report after the event yourself and share it in a venue that gives it a permanent Digital Object Identifier (DOI) [36] so it is easier for you to track citations.

Wherever your workshop information is shared, in whatever format, you have the chance to measure the impact of that information beyond the workshop's original remit. Effort is needed to track this impact through citations and other measures of views and use, for example, by using systems such as Altmetric [37], Google Analytics [38], YouTube, figshare [39], SlideShare [40], and Twitter Analytics [41]. This effort will help you to show the impact from your events and should form part of your overall measurement. These statistics should be tracked regularly, perhaps every six months or annually.

Another metric of impact beyond the room is whether participants talk about their experience positively with friends and colleagues. If you are running a workshop series, you could track recommendations by asking participants how they found out about the workshop. Referrals from previous participants is a good sign that you are doing something right.

## Conclusion

It is clear that you need to plan (Rule 1 and 2), use your knowledge and skills (Rule 3, 4, and 5), and apply techniques (Rules 6–10) to be able to measure the impact of workshops (those focused on exploring, learning, creating, or a mix). Ultimately, it is worth understanding why we want to measure impact in the first place and balance this with the amount of time required to organise the workshop and time we want to put into evaluating the workshop. With good measurements, we can convince funders to maintain and support the work that we do, encourage people to attend our workshops, and feel satisfied that the work that we are doing with our workshops is worthwhile and making a positive difference.

## References


1. Sufi S. How the Measuring the Impact of Workshops (MIW) meeting unfolded | Software Sustainability Institute [Internet]. 7 Oct 2016 [cited 19 Apr 2018]. Available from: https://www.software.ac.uk/blog/2016-10-07-how-measuring-impact-workshops-miw-meeting-unfolded
2. Pathways to Impact [Internet]. [cited 26 Apr 2018]. Available from: https://je-s.rcuk.ac.uk/Handbook/pages/GuidanceoncompletingaFellowshi/AccompanyingDocumentation/PathwaystoImpact.htm
3. Excellence with impact—UK Research and Innovation [Internet]. [cited 26 Apr 2018]. Available from: https://www.ukri.org/innovation/excellence-with-impact/
4. Higher Education Funding Council for England, Scottish Funding Council, Higher Education Funding Council for Wales, Department for the Economy. Initial decisions on the Research Excellence Framework 2021. 2017; 26.
5. Ponomarenko J, Garrido R, Guigó R. Ten Simple Rules on How to Organize a Scientific Retreat. PLoS Comput Biol. 2017; 13: e1005344. https://doi.org/10.1371/journal.pcbi.1005344 PMID: 28151954
6. Corpas M, Gehlenborg N, Janga SC, Bourne PE. Ten Simple Rules for Organizing a Scientific Meeting. PLoS Comput Biol. 2008; 4: e1000080. https://doi.org/10.1371/journal.pcbi.1000080 PMID: 18584020
7. McInerny GJ. Ten Simple Rules for Curating and Facilitating Small Workshops. PLoS Comput Biol. 2016; 12: e1004745. https://doi.org/10.1371/journal.pcbi.1004745 PMID: 27441642
8. Pavelin K, Pundir S, Cham JA. Ten Simple Rules for Running Interactive Workshops. PLoS Comput Biol. 2014; 10: e1003485. https://doi.org/10.1371/journal.pcbi.1003485 PMID: 24586135
9. Budd A, Dinkel H, Corpas M, Fuller JC, Rubinat L, Devos DP, et al. Ten Simple Rules for Organizing an Unconference. PLoS Comput Biol. 2015; 11: e1003905. https://doi.org/10.1371/journal.pcbi.1003905 PMID: 25633715
10. Workshops | Software Sustainability Institute [Internet]. [cited 19 Apr 2018]. Available from: https://www.software.ac.uk/workshops
11. Wilson G. Software Carpentry: lessons learned. F1000Research. 2016; https://doi.org/10.12688/f1000research.3-62.v2
12. Teal TK, Cranston KA, Lapp H, White E, Wilson G, Ram K, et al. Data Carpentry: Workshops to Increase Data Literacy for Researchers. International Journal of Digital Curation. 2015; 10: 135–143. https://doi.org/10.2218/ijdc.v10i1.351
13. Hackathon [Internet]. Wikipedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=Hackathon&oldid=836128128
14. Davis LN, McCallon E. Planning, Conducting, Evaluating Workshops. A Practitioner's Guide to Adult Education. 1st ed. San Diego: University Associates; 1975.
15. Mills-Scofield D. It's Not Just Semantics: Managing Outcomes Vs. Outputs. Harvard Business Review. 26 Nov 2012. Available from: https://hbr.org/2012/11/its-not-just-semantics-managing-outcomes. Accessed 19 Apr 2018.
16. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's Four Levels of Training Evaluation. 1st ed. Alexandria: ATD Press; 2016.
17. Kirkpatrick's Four-Level Training Evaluation Model: Analyzing Training Effectiveness [Internet]. [cited 19 Apr 2018]. Available from: http://www.mindtools.com/pages/article/kirkpatrick.htm
18. Perfect is the enemy of good [Internet]. Wikipedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=Perfect_is_the_enemy_of_good&oldid=827666643
19. Espeland WN, Stevens ML. Commensuration as a Social Process. Annual Review of Sociology. 1998; 24: 313–343. https://doi.org/10.1146/annurev.soc.24.1.313
20. Hawthorne effect [Internet]. Wikipedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=Hawthorne_effect&oldid=831866712
21. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. Plast Reconstr Surg. 2010; 126: 619–625. https://doi.org/10.1097/PRS.0b013e3181de24bc PMID: 20679844
22. Bryman A. Social Research Methods. 5th ed. Oxford University Press; 2015.
23. Unconference [Internet]. Wikipedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=Unconference&oldid=832523977
24. Likert scale [Internet]. Wikipedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=Likert_scale&oldid=830260749
25. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006; 3: 77–101. https://doi.org/10.1191/1478088706qp063oa

26. OpenRefine Community. OpenRefine: A free, open source, powerful tool for working with messy data [Internet]. [cited 19 Apr 2018]. Available from: http://openrefine.org/index.html
27. Crouch S, Hong NC, Hettrick S, Jackson M, Pawlik A, Sufi S, et al. The Software Sustainability Institute: Changing Research Software Attitudes and Practices. Computing in Science Engineering. 2013; 15: 74–80. https://doi.org/10.1109/MCSE.2013.133
28. Letter to Myself. In: HI Toolbox [Internet]. [cited 19 Apr 2018]. Available from: http://toolbox.hyperisland.com/letter-to-myself
29. Terhi Nurmikko-Fuller. Evaluating Learning through Games with a Purpose [Internet]. SoftwareSaved; 2016. Available from: https://www.youtube.com/watch?v=NqhcSqbiWD4
30. Nunes BP, Nurmikko-Fuller T, Lopes GR, Siqueira SWM, Campos GHB d, Casanova MA. Treasure Explorers–A Game as a Diagnostic Assessment Tool. 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT). 2016. pp. 217–221. https://doi.org/10.1109/ICALT.2016.136
31. Education and Construct 2—Scirra.com [Internet]. [cited 19 Apr 2018]. Available from: https://www.scirra.com/education
32. Unity—Showcase—Gallery—Non-games. In: Unity [Internet]. [cited 19 Apr 2018]. Available from: https://unity3d.com/showcase/gallery/non-games
33. Ericsson KA, Krampe RT, Tesch-Romer C. The Role of Deliberate Practice in the Acquisition of Expert Performance. Psychological Review. 1993; 100: 363–406.
34. Baaden M, Delalande O, Ferey N, Pasquali S, Waldispühl J, Taly A. Ten simple rules to create a serious game, illustrated with examples from structural biology. PLoS Comput Biol. 2018; 14: e1005955. https://doi.org/10.1371/journal.pcbi.1005955 PMID: 29518072
35. Ekins S, Perlstein EO. Ten Simple Rules of Live Tweeting at Scientific Conferences. PLoS Comput Biol. 2014; 10: e1003789. https://doi.org/10.1371/journal.pcbi.1003789 PMID: 25144683
36. ISO 26324:2012—Information and documentation—Digital object identifier system [Internet]. [cited 19 Apr 2018]. Available from: https://www.iso.org/standard/43506.html
37. Altmetric. In: Altmetric [Internet]. [cited 19 Apr 2018]. Available from: https://www.altmetric.com/
38. Google Analytics [Internet]. Wikipedia. 2018. Available from: https://en.wikipedia.org/w/index.php?title=Google_Analytics&oldid=827025596
39. figshare—credit for all your research [Internet]. [cited 19 Apr 2018]. Available from: https://figshare.com/
40. SlideShare.net. In: www.slideshare.net [Internet]. [cited 19 Apr 2018]. Available from: https://www.slideshare.net
41. Twitter Analytics [Internet]. [cited 26 Apr 2018]. Available from: https://analytics.twitter.com/about

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules on How to Organize a Scientific Retreat

**Julia Ponomarenko[1,2]*, Romina Garrido[1,2], Roderic Guigó[1,2,3]**

**1** Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain, **2** Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain, **3** Institut Hospital del Mar d'Investigacions Mediques (IMIM), Barcelona, Catalonia, Spain

* julia.ponomarenko@crg.eu



**Citation:** Ponomarenko J, Garrido R, Guigó R (2017) Ten Simple Rules on How to Organize a Scientific Retreat. PLoS Comput Biol 13(2): e1005344. doi:10.1371/journal.pcbi.1005344

**Published:** February 2, 2017



**Funding:** We thank CRG and the AGAUR Catalan agency for funding the retreats. We also acknowledge support of the Spanish Ministry of Economy and Competitiveness, "Centro de Excelencia Severo Ochoa 2013–2017." The funders had no role in decision to publish or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Scientific retreats are an intrinsic part of the life of many institutes, departments, and groups. They depart from traditional, virtual [1], and unconventional conferences [2], workshops [3], and other types of scientific meetings [4] in that participants generally all know each other prior to the retreat, and, often, they have a good grasp of the scientific interests and accomplishments of each other; they may even be working closely together. Participants, thus, do not attend the retreat expecting to necessarily hear about breakthroughs in their fields of interests or to present their latest results to an expert audience but rather to have a deeper knowledge of the work of their closest colleagues, learn from developments in related areas, and explore potential collaborations.

Since retreats usually take place away from the home institute and may expand over two or more days, they are expensive to organize—including significant institutional funds and employees' working and personal time. They are disruptive of the daily scientific routine: experiments may need to be stopped or planned ahead, and regular activities such as seminars and group meetings need to be cancelled. Thus, to many, the benefits of moving a group of people who already share the same working space to a remote location over a period of two or more days are not obvious. After all, retreat participants already have the opportunity of meeting, almost on a daily basis, at the home institute.

There is little empirical evidence that scientific retreats lead to better science (whatever this exactly means); we have been unable to find any work that would correlate frequency or length of scientific retreats with any of the metrics usually employed to measure the quality of science. Yet, anecdotal evidence of a positive correlation between scientific breakthroughs and scientists being outside the lab is abundant and includes a discovery of penicillin attributed to a long summer vacation by Fleming in 1928 [5] or a discovery of Velcro by Georges de Mestral after a hunting trip with his dog in 1941 [6]. More recently, the invention of a new cipher for using DNA as high-capacity data storage by Ewan Birney and Nick Goldman of the European Bioinformatics Institute (EBI) apparently happened involving "many beers" [7, 8].

If properly planned, retreats offer an informal environment, which is becoming increasingly rare with the "laborization" of science, when scientists tend to follow preestablished working schedules and interact with each other only during the regular working hours, following well-structured formats of group meetings, conference calls, seminars, and other meetings. There is also an increasing divide between work at the lab and personal life. These tendencies are new to science, often being seen in the past as a way of life rather than a means of living. Retreats offer the possibility to break these tendencies—even if only for a short period of time—by bringing together work and personal life. At the retreat, a student may have a lunch with a professor he never had a chance to interact with, postdocs from different groups may hike and party together, and principal investigators (PIs) with little time for informal discussions during a normal working week may have a chance to debate about their favorite topics and get to know each other. All these apparently irrelevant events may result in productive science. Smaller retreats—involving, for example, a single research group—can provide an opportunity for improving the day-to-day group dynamics [9] and boosting creativity [10]. Retreats can also facilitate building trust amongst people that is vital for productive working relationships [11]. The authors believe that the intangible benefits delivered by scientific retreats are currently vastly underrated and understudied.

Retreats may come in many flavors. They may appeal to a particular constituency (PhD students, postdocs, PIs), and they may involve hundreds of participants from an entire institute or barely a dozen from a single research group. They may be attended only by scientists or they may also include administrative staff. For the past 11 years, the authors have organized a yearly two-day retreat for the Bioinformatics and Genomics Program at the Center for Genome Regulation (CRG) in Barcelona, Catalonia. This event also includes computational biology groups from other programs at CRG as well as other institutions from the Barcelona Biomedical Research Park (PRBB). It could be described as a departmental retreat, attended by students, postdocs, PIs, technicians, and administrative staff. The 2016 retreat took place in a rural mountain hotel an hour's driving distance from Barcelona. In previous years, we opted for a beach location or a ski resort, always no more than two hours from the city. In 2016, there were more than 110 scientists present, the majority of whom were PhD students and postdocs. According to the feedback we have received from the participants, the retreat provided a memorable and for some participants even a life-changing experience.

With scarcity of information on how to organize a scientific retreat, organizers often opt to follow a standard formula of a conference with an extensive scientific program. And while retreats and other scientific events share common rules, many of which can be found in the *PLOS Computational Biology* "Ten Simple Rules" collection [1–4], here we emphasize rules applied specifically to retreats: in particular, departmental-style retreats.

## Rule 1: Define the Purpose

Decide first on why (and whether) you need the retreat. Engage participants in deciding about the purpose and clearly communicate the purpose to participants. Depending on the type of the retreat and the participants, the answers might be found in the following statements: "Discuss the future of the field," "Discuss career perspectives," "Learn what colleagues are working on," "Get to know your colleagues," "Foster the feeling of belonging and bonding," "Make friends," and "Relax and have fun." A pre-retreat survey can help to learn about the participants' expectations and needs. Our latest pre-retreat survey is provided in S1 Text.

By launching a pre-retreat survey, which was answered by 99 participants of the previous retreat, we learned that we should focus on the three major goals: "Exchange information on who is working on what with the goal to foster new collaborations," "Get to know each other better," and "Have fun." Those goals were logical for our retreat because it gathered scientists from five institutions at the PRBB campus, many of whom were students and postdocs of multiple affiliations. In the end, we defined the goal of the retreat as "Fostering the feeling of belonging in our shared computational biology community."

## Rule 2: Define the Budget, Length, and Time of the Retreat

The length and time of the retreat are substantially defined by the available budget.

Although a two-day retreat might be of the ideal length—since spending one night out gives much more opportunities for interaction—if the budget is limited and the number of participants is small, a one-day retreat can be justifiable, with dinner or just going out to the bar at the end of the day. A retreat is an intense event and extending it to more than two days may be exhausting.

It should be expected that the retreat cost can considerably increase at summer time, school breaks, and holiday seasons; also, depending on personal circumstances, it may be complicated to organize it during a weekend. Therefore, we recommend organizing a retreat at the beginning or at the end of a working week. Considering numerous events happening during a year (symposiums, training, classes, attestations, seminars, thesis defenses, etc.), it is extremely important to schedule a retreat a year in advance.

From our experience, we suggest budgeting or taking care of the following on your own:

- transportation to and from the venue by bus, boat, train, etc.
- lodging
- meals: breakfasts, coffee breaks, lunches, dinner(s)
- meeting rooms and equipment (microphones, projectors, computers, boards, laser pointers, etc.)
- name and/or team badges or stickers
- equipment and/or help for outdoor activities (e.g., volleyball equipment, local hiking guide, etc.)
- room and/or place for a party and equipment (karaoke, music, projector, etc.)
- music playlist or a DJ
- bar (define and communicate drinking policies in advance: who pays for drinks, how much they cost, open and closing times of the bar, etc.)

## Rule 3: Decide on the Venue

Depending on the retreat goals, budget, weather, and the number, age, and seniority of participants, the venue has to be carefully selected to provide opportunities for all planned activities and productive encounters. The venue is hugely important for the retreat success since, in contrast to a conference, people cannot escape. They have to be physically present for the whole duration of the event in the same environment where they will be sleeping, eating, resting, talking, drinking, etc.

For the most recent retreat, we selected an isolated rural hotel that provided two large conference rooms, a large dining room, various cozy and private places for encounters and discussions, as well as plenty of opportunities for outdoor activities, such as hiking, football, an adventure park, and a swimming pool. There was also an indoor entertainment hall with a bar for parties, dancing, and karaoke.

Transportation to and from the venue needs to also be considered. As a rule, for each retreat, CRG provides a coach from and to the work place. But since some participants live in different parts of and outside Barcelona, upon registration we ask participants if they will be using a coach or not. Before the retreat, the lists of those who will use a coach are made for checking in everyone on the way to the bus. A car-sharing signup option can also be considered.

## Rule 4: Establish Policies

It is your responsibility as a retreat organizer to ensure the participants' safety and appropriate behavior during the retreat to avoid any kind of harassment, including sexual [12]. If your organization has policies written specifically for retreats and/or general rules of conduct, we suggest reminding participants about these policies at the time of registration, emphasizing that the retreat is an extension of a work environment and all rules of conduct have to be respected. However, if there are no such policies in place or they need to be changed, use the upcoming retreat as an opportunity to address this issue.

Depending on the local customs and law, you might consider outlining media policies regarding using mobile phones, taking pictures, and recording during the retreat. To promote the feeling of trust and to allow people to relax, you might also decide on a rule about sharing the retreat collaterals (see Rule 10).

Do not overlook the gender balance issue [13], considering both the scientific and social programs and while forming the retreat committees.

## Rule 5: Set Up the Agenda and Appoint Committees with Responsibilities and Timelines

The program should include time for scientific and team-building activities, food and entertainment blocks, and plenty of free time. Generous timing should also account for any unforeseen issues and hiccups, such as bad weather, talk cancellations, broken equipment, etc. Do not also overlook local customs concerning meal and bedtime; for example, in Spain, lunch at very early afternoon hours, dinner before 8–9 pm, or a party stopped at midnight would be considered unusual. The agenda of our 2016 retreat is provided in S2 Text.

We can recommend setting up an organizing committee including one secretary or administrator and one or two senior scientists or professors. The organizing committee should take care of the logistics, finances, and general coordination of the retreat. Our experience also shows that at least two additional committees are helpful to form: a scientific committee that will take care of the agenda, scientific sessions, and other scientific events, and a social committee that will organize social activities.

A kickoff meeting defining the main goals, expectations, responsibilities, and timelines of all committees should be organized at least three months in advance. Afterwards, committees can work independently. The final meeting should take place a few days before the retreat.

## Rule 6: Arrange an Engaging Science Program

It can be regrettable to spend precious retreat time on presentations that can be delivered via seminars, meetings, data clubs, and other regular institutional gatherings. We recommend the retreat for probing unusual formats, some of which have been employed by "unconferences" [9]. Think of, for example, organizing a few topic-centered sessions of short talks delivered by both faculty and students, followed with a panel discussion. Depending on the participants and the size of the retreat, it might be useful to include tutorials, short workshops, and roundtable discussions. Debates and games, such as scientific quizzes, scientific "speed dating" [14], and "fishbowl" [15] are excellent and engaging forms to discuss various topics. As quizzes were a staple of previous retreats, this year we decided to organize Oxford-style debates, with two debating teams and a moderator. Run at the climactic time of the retreat before the first day's dinner, two debates—one discussing what a gene is and the other the cons and pros of pre- versus post-publishing reviews and the future of scientific publications—provided enough fuel for the buzz to continue during dinner.

From our survey, it also became clear that participants wanted to hear from PIs about their careers and about the future of the field. We therefore asked three PIs who are about to finish their junior tenure and move to other institutions to each give an hour talk reflecting on their careers and their views on the science. These talks were engaging and moving for both PIs and audience; they also created an intimate and special atmosphere for the whole retreat. Indeed, it has been shown in multiple studies that memories are more readily made and logical reasoning performance is improved when positive emotions are stirred [16, 17].

## Rule 7: Organize a Range of Social Events

Without outdoor activities or time to relax, nobody would feel that they are at a "retreat." Do not however rely only on self-organization by just allocating time in the program for these activities. Plan the activities sensibly: for example, if you schedule in the morning of the second day time for volleyball, nobody will show up to play; likewise, free time for hiking after lunch will likely turn out to be a time for a swim or a nap. Instead, explicitly organize every hour of activities, accommodating various interests and physical abilities. The latter cannot be stressed enough as you cannot force everyone to participate in extreme physical activities that may as well involve humiliating moments.

In the pre-retreat survey, we asked people about outdoor and team-building activities they wanted to suggest and would be interested in participating and/or organizing. After selecting a few activities and forming organizing teams, we asked participants to register for as many activities as they wished, specifying a ranking of choice. In the result, we had two hikes of different difficulty levels, a 10-km run competition, a football competition, a capoeira session, and a tree-climbing activity in an adventure park. The majority participated in at least one of these activities, while some still opted for a couch, a swimming pool, or a good read. Most important was that everyone could return from the retreat to work with renewed energy and motivation.

The retreat is not a retreat without a proper planned after-dinner party. Why is partying important? It is time to celebrate excitement for hard work and to rejuvenate the team spirit. It allows people to bond with and to show gratitude towards each other. It fosters camaraderie, since it is easier to see yourself as a team when you are enjoying yourselves together. It allows improving verbal and nonverbal communication skills. To avoid embarrassments and tiredness associated with excessive drinking, do not rely only on a DJ, a bar, and self-organization. Get creative: think about having science-related themes for the party. For some retreats, we organized scientific competitions, games, and quizzes. This year we organized a team karaoke "Eurovision-like" competition with awards and randomly formed teams. We suggest involving participants in the organization of the entertainment. Engage a good dancer to encourage people to dance in formation, which provides a tangible bonding experience. Ensure that the place has proper equipment and space for a party. Have your most party-inclined members take the lead, or, if the budget allows, consider hiring a professional entertainer to help to get the party going after a long retreat day.

## Rule 8: Ensure Everyone Is Included

The retreat is about its participants; and to make it engaging, productive, and memorable for everyone, it must include enough opportunity for each individual to speak and to work on their "soft skills." Activities organized by committees formed by participants, such as a debate, a game, a sport, or a karaoke competition, give an excellent opportunity for everyone to tap into hidden potential and train hugely important team-building and communication skills.

Diversity of personality types or traits [18, 19] and team roles [20–22] should be also taken into account while planning talks, debates, and other activities so that strengths and competences will be matched to particular roles. For example, you might think whether randomly formed teams will work for a particular activity or if it is valuable to organize teams in advance.

Here is an example of how we used the format of short talks and an informal lunch as team-building and public speaking exercises. Sessions of two-minute lightning talks were regularly run at this retreat and were loved by speakers but not so much by the audience, who were lost in the ocean of talks. This time, we decided to have a single one-hour session of three-minute teaser talks using the format: "The scientific problem that twists my brain and I want to discuss during lunch is. . ." Individuals were randomly assigned to teams of six. The number of talks corresponded to the number of teams, and each team was assigned a table to discuss the lightning talk during the following lunch break. Knowing about this format in advance, some presenters chose to raise topics requiring brainstorming—for example, "How to start a PhD thesis without getting lost" or "From circular to linear: How to represent gene order?" This approach of discussing science around a very specific problem during lunch proved to be productive and fun.

## Rule 9: Plan Ahead and Keep to Time

There are many large and small tasks to take care of well in advance of the retreat, including registration, pre-retreat surveys, transportation, dietary and/or accessibility requirements, scheduling of each event, name badges, equipment, and many others. If not planned in advance, some things may take an unexpectedly long time and cause a disruption in the schedule: take, for example, the logistics of checking in 100 people at the hotel.

The importance of good time keeping cannot be stressed enough, and importantly, it concerns not only talks and mealtimes but all other events as well. Remember to wrap up the event with a conclusion, accolades, and awards (see Rule 9 of Corpas et al. [4]).

Have an efficient and experienced administrator to manage the retreat logistics. It must be a rule for organizing any event: if you cannot count on an excellent administrator, you are doomed to fail. For more on what an administrator might and should delegate, see Rule 6 of Corpas et al. [4].

The success of the retreat also depends on the level of participants' involvement and commitment; so ideally, set up a dedicated committee for each program area.

## Rule 10: Gather and Give Feedback

Sharing the retreat collaterals—presentations, photos, and videos—will make it more memorable. At our organization, Dropbox is commonly used to share folders and files; therefore we set up a Dropbox folder shared with all retreat participants. Other organizations might opt for a different platform. What is most important is creating shared memories and experiences that are key factors in team building and producing good leaders in your organization and/or community [23].

Of course, there cannot be a perfect retreat, so gathering feedback on what worked and what did not can help planning for the next time. We usually gather feedback about retreats, courses, and other events via online surveys; in our experience, a survey conducted immediately after an event engages up to 80%–90% participants. We also suggest obtaining feedback on new ideas and proposed format changes a few months before planning the next retreat via a pre-retreat survey.

At the retreat, an administration and group leaders can also learn a lot about their leadership style and which type of leaders they have in their groups [24–26], as well as what kinds of personalities are needed for the group to be more productive, creative, and stress-free [27].

## Supporting Information

**S1 Text. Pre-retreat survey of 2016.**
(DOCX)

**S2 Text. 2016 retreat agenda.**
(DOCX)

## Acknowledgments

We thank the PRBB Computational Genomics community for making the retreat possible.

## References


1. Gichora NN, Fatumo SA, Ngara MV, Chelbat N, Ramdayal K, et al. Ten simple rules for organizing a virtual conference—anywhere. PLoS Comput Biol. 2010; 6: e1000650. doi: 10.1371/journal.pcbi.1000650 PMID: 20195548
2. Budd A, Dinkel H, Corpas M, Fuller JC, Rubinat L, Devos DP, et al. Ten Simple Rules for Organizing an Unconference. PLoS Comput Biol. 2015; 11(1): e1003905. doi: 10.1371/journal.pcbi.1003905 PMID: 25633715
3. Pavelin K, Pundir S, Cham JA. Ten simple rules for running interactive workshops. PLoS Comput Biol. 2014; 10: e1003485. doi: 10.1371/journal.pcbi.1003485 PMID: 24586135
4. Corpas M, Gehlenborg N, Janga SC, Bourne PE. Ten Simple Rules for Organizing a Scientific Meeting. PLoS Comput Biol. 2008; 4(6): e1000080. doi: 10.1371/journal.pcbi.1000080 PMID: 18584020
5. Haven K. Marvels of Science: 50 Fascinating 5-Minute Reads. 1st ed. Littleton, Colo: Libraries Unlimited; 1994.
6. Stephens T. How a Swiss invention hooked the world. 2007. http://www.swissinfo.ch/eng/how-a-swiss-invention-hooked-the-world/5653568
7. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, Birney E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature. 2013; 494(7435): 77–80. doi: 10.1038/nature11875 PMID: 23354052
8. Yong E. DNA storage: The code that could save civilisation. BBC, 2013. http://www.bbc.com/future/story/20130724-saving-civilisation-in-one-room.
9. Tuckman BW. Developmental sequence in small groups. Psychological Bulletin. 1965; 63 (6): 384–399.
10. Rickards T, Moger S. Creative leadership processes in project team development: An alternative to Tuckman's stage model. British Journal of Management 2000; 11(4): 273–283.
11. Engelbrecht AS, Heine G, Mahembe,B. The influence of ethical leadership on trust and work engagement: An exploratory study. SA Journal of Industrial Psychology. 2014; 40(1): 1–9.
12. Clancy KB, Nelson RG, Rutherford JN, Hinde K. Survey of academic field experiences (SAFE): trainees report harassment and assault. PLoS ONE. 2014; 9(7): e102172. doi: 10.1371/journal.pone.0102172 PMID: 25028932
13. Martin JL. Ten Simple Rules to Achieve Conference Speaker Gender Balance. PLoS Comput Biol. 2014; 10(11): e1003903. doi: 10.1371/journal.pcbi.1003903 PMID: 25411977
14. Lawson J. Getting together: Science speed dating. 2013. Nature Blog. http://www.nature.com/scitable/blog/conferencecast/getting_together_science_speed_dating.
15. Fishbowl: The art of active listening. http://slitoolkit.ohchr.org/data/downloads/fishbowl.pdf.
16. Jung N, Wranke C, Hamburger K, Knauff M. How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. Front Psychol. 2014; 5:570. doi: 10.3389/fpsyg.2014.00570 PMID: 24959160
17. Kensinger EA. Remembering the Details: Effects of Emotion. Emot Rev. 2009; 1(2):99–113. doi: 10.1177/1754073908100432 PMID: 19421427
18. The Myers and Briggs Foundation. The MBTI Personality Types. http://www.myersbriggs.org/my-mbti-personality-type.
19. Wikipedia. Fundamental interpersonal relations orientation. https://en.wikipedia.org/wiki/Fundamental_interpersonal_relations_orientation.
20. Belbin M. Management Teams. London; Heinemann; 1981.
21. Belbin M. Team Roles in a Nutshell. http://www.belbin.com/media/1335/belbin-for-lecturers.pdf
22. Belbin Team Roles. http://www.belbin.com/about/belbin-team-roles/

23. Maxwell JC, Parrott L. 25 Ways to Win with People: How to Make Others Feel Like a Million Bucks. Thomas Nelson; 2005.

24. Goleman D. Leadership that Gets Results. Harvard Business Review. 2000. https://hbr.org/2000/03/leadership-that-gets-results.

25. Kets de Vries MFR. The Eight Archetypes of Leadership. Harvard Business Review. 2013. https://hbr.org/2013/12/the-eight-archetypes-of-leadership

26. Antonakis J, House RJ. Instrumental leadership: Measurement and extension of transformational–transactional leadership theory. The Leadership Quarterly. 2014; 25 (4): 746.

27. Driskell JE, et al. What makes a good team player? Personality and team effectiveness. Group Dynamics: Theory, Research, and Practice. 2006; 10.4: 249.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Curating and Facilitating Small Workshops

**Greg J. McInerny**[1,2]*

**1** Centre for Interdisciplinary Methodologies, Department of Social Sciences, University of Warwick, Coventry, United Kingdom, **2** Department of Computer Science, University of Oxford, Oxford, United Kingdom

* g.mcinerny@warwick.ac.uk

CrossMark
click for updates

OPEN ACCESS

**Citation:** McInerny GJ (2016) Ten Simple Rules for Curating and Facilitating Small Workshops. PLoS Comput Biol 12(7): e1004745. doi:10.1371/journal.pcbi.1004745

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** July 21, 2016



**Funding:** GJM was supported by the 2020 Science programme, funded by the EPSRC Cross-Disciplinary Interface Programme (grant number EP/I017909/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

As a participant, workshops are by far my favorite scientific event. Compared to conferences, the interactions can be more intense, discussions can be deeper, and the resulting collaborations are often stronger. Working with 10–30 attendees over a few days can lead to a more open and integrated event than a conference. At workshops, you are a participant in the whole event, and you can make many direct contributions to its goals. In contrast, at conferences, the aim is for a broad informational in which you are part of the audience and contribute comparatively little content.

As an organizer, workshops will present you with diverse challenges. You will manage the project and its logistics [1], and you will also be the curator(s) by developing and negotiating workshop content. Possibly the least well recognized role is as facilitator(s), when you enable interactions amongst participants and workshop activities. In addition to organizational skills [1], a workshop will profit from your creativity, empathy, and mediation skills. These ten simple rules make links between these roles and aim to help you reach your goals whilst making an enduring contribution to your community [2].

There is no single formula for creating good workshops. In contrast to the fairly standard format of conferences (plenary, coffee, talks, lunch, talks, coffee, talks—repeat), workshops can take diverse forms, and indeed, they should to fit different goals. For example, different workshops are needed when exploring a single research topic, initiating a working group, developing interdisciplinary collaborations, or testing new methods and software (e.g., compare [1] to [3]). Different workshop goals will then require different kinds of attendees, timetables, interactions, props, atmospheres, etc. The details of your roles as organizer(s), curator(s), and facilitator(s) will also differ between different types of workshops and will develop as your experience and confidence grows.

Developing workshops can involve jeopardy. The organization, curation, and facilitation can, in places, go wrong. Participants' time away from their work and personal lives should be worthwhile (also see [4]). Tangible outputs can also be hard to develop in a short period and may not always involve you. Most obviously, the workshop will divert time you would spend on research and other parts of your job. It is perhaps the wrong framing to see workshop organization as a "time- and energy-draining black hole" [5]. Like all worthwhile things, workshops will require your time and energy, but that needn't be draining. Instead, you can make it a rewarding and energizing experience. Your dedication and enthusiasm will reduce the jeopardy and increase the productivity of the workshop. You will find greater enjoyment in the whole process, too.

## Rule 1: Assess Past Successes and Failures

Some workshops are better than others. But why? Was it the organizers, attendees (see Rule 3), goals, size, subject matter, timetabling, activities, venue, location, seating plan, seats, timetable, time keeping, props, access to plug sockets, projection equipment, biscuits, heating, travel, weather? Individually, any of these can be a minor factor. Some, like appropriate attendees and clear goals, will directly influence the success or failure of your workshop. Take control of what you can, and make the event what you want it to be. Think big and small.

Make workshops part of your conversations. Find out what people like or have found irritating. Explore your contacts for organizers of workshops and events. Find out what succeeded and why. You might ask them to be a co-organizer. Workshops should vary, so don't follow someone else's template uncritically. Most of all though, be clear about what you want to achieve. Is a workshop really the best process to achieve that goal?

## Rule 2: Develop a Brand

Branding helps you engage with people throughout the development, execution, and reporting of your workshop. A brand should efficiently explain what you will do and why (e.g., a catchy workshop name, a recognizable logo, and a tweetable mission statement or statement of goals). Successful brands needn't require huge investments of time, and brands needn't be loud to be noticed. Aim for "salient" rather than "extravagant" (Fig 1).

Ignoring branding increases your chances of miscommunications and missed opportunities. Your brand will support you with: (a) pitch (securing funding, inviting participants, informing speakers and co-organizers), (b) function (developing identity and focus, producing printed materials and signage, a succinct reminder of the workshop goals), (c) recognition (reporting, gaining attention and reputation, developing buy-in and helping people find the website), and (d) decision making (a frame of reference for both organizers and participants). Brands are just as important, if not more, for single workshop events where outcomes (however small) must be produced in a short time. Also, your invitations may have a less substantial reputation to rely on and must stand out amongst busy email inboxes.

Importantly, you will be the most visible and vocal component of your brand, the one thing that interfaces with everything else. Be prepared to put yourself out there and represent and bring coherence to the workshop.

## Rule 3: Recognize Diversity and Use It

Workshops become conferences if they are too big, have too many talks, or if you invite an audience rather than participants. Workshops become group meetings if too many people


**Fig 1. Branding needn't be demanding.** The "Death of the Desktop" workshop (see: http://beyond.wallviz.dk/, https://vimeo.com/102527731) had a strong brand that was intriguing, evocative, and memorable. The twitter account (@visfutures) revealed information and contributions in the lead-up to the event. The brand helped to set a tone for the preworkshop submissions and the workshop activities. Consider how effective the brand would have been if the workshop was just called "Consequences of a concerted community focus on visualization technologies and the impact on interaction and display techniques." Would this have been easy to remember or Google? Would this have set a theme that was easy to follow?

doi:10.1371/journal.pcbi.1004745.g001

attend from one organization or project, become training events if there are too many early-career researchers, and become board meetings if there are too few. Boring workshops emerge from too much agreement, too little participation, or too little openness. Workshops depend on diversity, so be prepared to be purposeful with inviting participants, and don't be afraid to engage in a bit of social engineering (also see [3]). Some workshop dynamics can be hard to foresee, but others are very predictable.

Invite a diversity of characters, skills, and knowledge that will make your workshop work. Invite provocative characters that can ignite discussions or will be likely to play devil's advocate, and invite those who can provide syntheses to multifaceted discussions. Invite potential authors for paper outputs. Why not invite the editor of a journal? The right attendees are fundamental to achieving your goals.

Group activities require rapporteurs, facilitators, and contributors as well as skills such as drawing. If groups self-select and deliberate on task management, you will waste time, de-energize participants, and disrupt the flow. So, you can assign groups (and seating plans, Rule 6) with purpose. These groupings should ensure function. For instance, if vocal individuals will dominate discussions, then place early-career researchers or quieter individuals in their own group. Use specialist groups to hear discipline-specific views, and use multidisciplinary groups to understand how to integrate different perspectives. Create structure from the diversity.

## Rule 4: Tell Speakers What to Say

Organizers, participants, and speakers all appreciate when talks are relevant, so offer appropriate direction so that presentations fit the workshop goals. This will require some negotiation and gentle editorial input. If you want a particular presentation, you will have to ask! Without direction, speakers can unintentionally divert a workshop's path. Provide context on the goals and the backgrounds of participants beforehand. Some speakers really enjoy new challenges, so you might ask for something nonstandard; for example, a view from the shop floor or an outsider's view on a topic. It may just change the wording of an existing talk rather than the whole presentation. Either way, all speakers like to be relevant and on point.

Not everyone will adjust their talk based on overlap with preceding talks. So, where possible, reduce the potential for awkwardness and repetition by breaking down topics into modules. You might encourage speakers to consult with each other, or consider sharing your introduction so they can edit overlapping material. This can help direct your own presentation (also see [6]).

## Rule 5: Learn to Love Emailing

Emails are the backbone of workshop preparations (invites, replies, accommodation, replies, requests, replies, travel plans, replies. . .phew!). It can be tempting to hand off many of these tasks (hundreds of emails for a workshop? It has been known!). However, emails are opportunities to build relationships and uncover opportunities. You may have the support of event organizers, but work with them rather than being dependent on them.

Emails don't have to be too formal and should encourage a dialogue. Emails can help you understand participants' backgrounds, motivations, and expectations that can then reveal any concerns and suggestions they may have and may also help you consider seating arrangements (Rule 6). Not every email will be interesting. Most won't. Taking responsibility means information is not lost on the way to someone else's spreadsheet. You can tailor invitations to different kinds of people and different disciplines ("you would really find X, Y and Z of interest"), offer help before being asked (e.g., travelling with a partner = double room, or being family-friendly [7]), address concerns at an early stage ("thanks for the invite but I don't think I am the right person. . ."), and pick up opportunities ("our organization would like to sponsor the workshop").

Whilst many tools can make mundane tasks easy (e.g., SurveyMonkey, Doodle, Google Forms), consider when group requests for information are suitable or when individual emails may be better. A poorly pitched group email can spark off many individual responses. Be prompt in your own replies, as participants and event administrators will need information at certain times.

## Rule 6: Seriously, Consider Your Seating Plan Seriously!

Conference-style, row seating plans can stifle productive workshop dynamics. A coat laid on a seat, people occupying row ends, and a lack of eye contact can all inhibit important introductions and exchanges. Seating plans should account for practical matters (view of projector screens, fire exits, position of plug sockets, or access to workshop materials and props) and support the flow of your timetable and the interactions amongst participants. It is actually a big deal, so take it seriously! Also, see [3].

"Cabaret" seating plans can help workshops to function (Fig 2). People can maintain views of screens and speakers whilst allowing face-to-face contact in discussions. In addition,

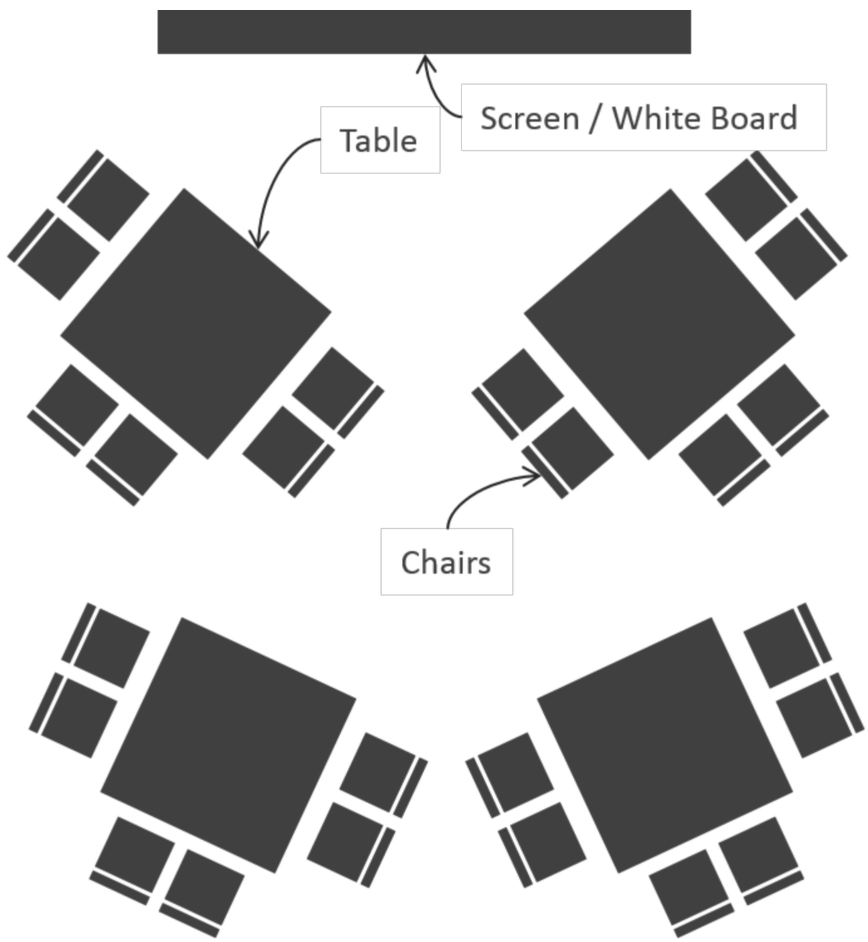



**Fig 2. Cabaret seating plans (also see [3]) can support different workshop activities.** A seating plan may help you distribute facilitators and rapporteurs (See Rule 3).

doi:10.1371/journal.pcbi.1004745.g002

switching between whole and smaller group activities is made easy. A named seating plan can reduce uneasiness and help direct different groupings for different activities or days (if you know people and their work sufficiently well). A spare chair or two at the back may allow people to find a bit of thinking space. Think ahead and don't be afraid to ask venues for what you want.

## Rule 7: Document the Workshop Using Fresh Resources

There are many opportunities to go further than traditional recordings or note taking. Twitter can help you in a variety of ways [8], including recording key points, links, names, and organizations in a format that is immediately shareable. You can also engage audiences beyond the workshop and solicit their views or questions. Active tweeters make this easier, so invite some! Leave some tweets before the event as breadcrumbs and also tweet a link to the workshop website with an event hashtag. However, tweets may not always be essential or appropriate to the workshop material and sometimes can even be a distraction.

Figshare, SlideShare, and related repositories make it easy to disseminate and discover talks if they are tagged appropriately. Tweeting links to talks can generate extra interest and value for your participants. Relevant or selected tweets and documents can also be recorded (e.g., storify.com) or retweeted via a workshop twitter account. Not everything has to be shared though, and you might consider a charter that guides the dissemination of material beyond the workshop.

At one workshop, a very talented participant showed me their sketchnotes (Fig 3)—"purposeful doodling" (http://sketchnotearmy.com/about/)—that record selected content from talks or discussions. Sketchnotes can be more evocative and memorable, capturing key points or images or adding context by drawing themes together. Find a sketchnoter and invite them along. Brief them on the workshop beforehand.

## Rule 8: Include Something Unexpected

Executing a standard workshop format is demanding and worth trying in the first place [9]. When your confidence has grown, you may consider more inventive activities that break up

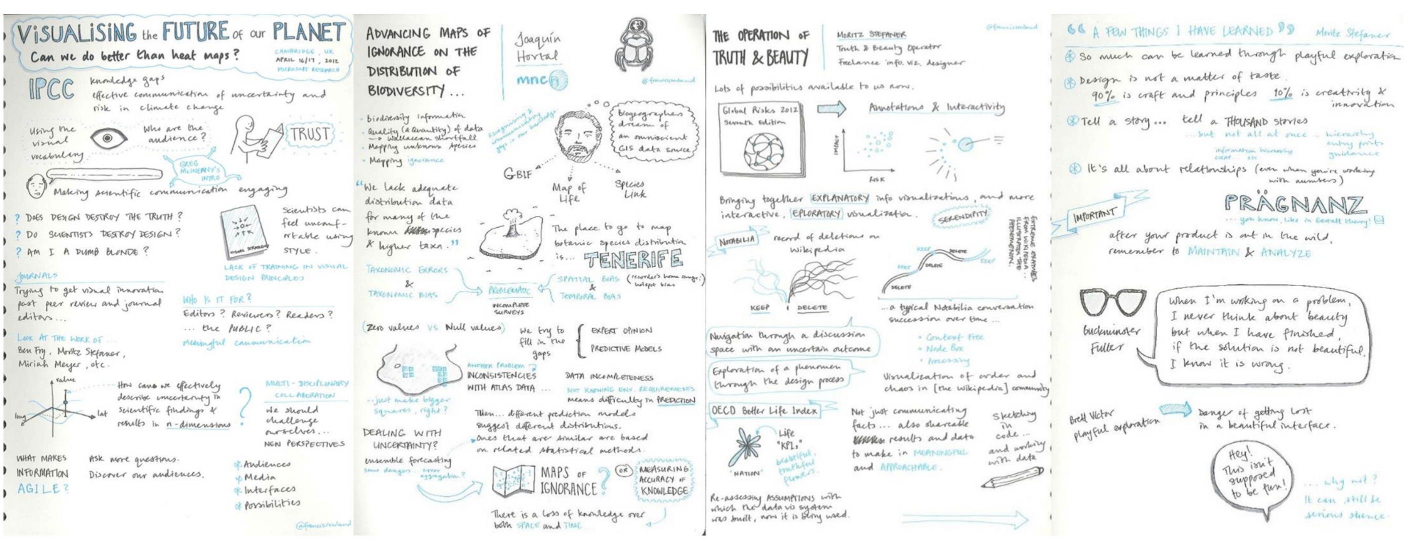



**Fig 3. Sketchnotes (http://www.flickr.com/photos/francisrowland/6944419112/lightbox/ | @francisrowland).**

doi:10.1371/journal.pcbi.1004745.g003

routine interactions and allow you to go in different directions. This could be simple, from inviting an unexpected speaker to a more involved approach with structured activities, such as "gamestorming" (http://www.gogamestorm.com/).

Used critically, gamestorming can help you dissect, characterize, and explore ideas or problems in ways that discussions cannot. By harnessing participants' creativity and skills other than talking from a crowd, these "games" can produce much more inspired responses to problems by breaking down preconceived ideas and building new perspectives. Participants can find flaws in these simple games, but that is often useful in itself. It can also be "fun."

Unexpected speakers can help explore the extremes of topics. It's a simple and effective strategy. They may be people who use different tools for similar research, similar tools for different research, or are just somehow unusual in their field. Deliberately stretching a topic's boundaries can stretch the workshop's possibilities. At a scientific workshop, invite an information visualization speaker; at a visualization workshop, invite a social scientist; and at a multidisciplinary workshop, invite a communications expert. Your job is to ensure that this strategy brings value and isn't just a diversion.

Remember that there aren't really any rules. Presentations don't have to have equal time. If topics are familiar, then allocate less time; if they're unfamiliar, add more time for both presentation and questions. You can pair talks and leave questions for a discussion afterwards, use the popular "ignite talks" formula, have a demo session over coffee, use Pecha Kucha (http://en.wikipedia.org/wiki/PechaKucha), incorporate an "unconference" component in which some activities are crowdsourced [10], or go for a hike together. Or do all of those. Be inventive. You don't have to do what everyone else has done before. Overly complicated timetables can, however, become bewildering, and hikes may require wellies, so be sensitive to your participants.

## Rule 9: Prepare for the End

Final impressions are important. How you close a workshop can affect how participants perceive a workshop's success and how they judge their own contributions. So, invest in the workshop's closing [1]. It should match the vibrancy of your event. If you finish early, what does that say to the participants?

Prepare a structure for your final words and any thank-yous. The gaps can be filled in during the workshop; e.g., record take-home points from talks and activities, key suggestions, solutions, or any extra thanks that should be noted. Do not make too many unplanned promises though! Many workshop papers are never written. Gauge support and seek informal feedback before launching into public requests for significant work after the event. This avoids the tumbleweed (silence) when asking who would like to lead on a particular output.

At the end, you and the participants will be at your most exhausted and gorged on information and starting to think of return journeys, work, and families. Asking participants for a take-home message can be awkward and ineffective at this point. Even with only 20 or so people, tired, on-the-spot reflections can provide a damp ending. You can brief and plant selected rapporteurs beforehand if needs be (use diversity strategically, Rule 3).

## Rule 10: Don't Panic, Things Go Wrong

As with the rest of life, things can go wrong. It's how you respond that matters. Don't panic, and deal with problems with a smile. Many of these simple rules have emerged from three types of problems.

First, problems with organization and planning: food may not arrive, dietary requirements may not have been passed on, and speakers may forget your previous discussions (Rule 4). A venue might inform you that it has removed projection equipment, Wi-Fi, flipcharts, furniture, and all facilities support a week before (it has happened!). Do not look for fault or blame. Just focus on solutions, and act quickly. Double check details close to the actual date. If possible, visit venues beforehand and check seating or even taste food if that is possible. Not all spring rolls are created equal.

Second, problems with not sticking to your plan: unexpected discussions or opportunities can tempt you to deviate from your timetable. Groups may not have completed a task on time. Or, you may panic about an activity's relevance and ditch it. Improvising can trip up the workshop flow or put pressure on people and activities later in the schedule, so be very sure that the alternative is better and be decisive. You don't have to accommodate on-the-spot requests to change the timetable or talk about a particular subject, even if the request comes from someone more senior. Sticking to a well-thought-out plan is often the best response. A good discussion will find its own way of continuing, so don't worry about exhaustively working through every item to a conclusion.

Thirdly, unexpected problems: some things just happen and you can't do anything about it. Don't panic. Do what you can and then go with the flow. The worst cases are illness during workshops and mislaid plane tickets. You can't do much except making sure the participants' needs come before those of the workshop. It is usually a good story after the event.

## Summary

Rather than overcomplicating your workshop plans, you need to strike a balance amongst what things would be interesting to do and those that have to be done. It can be easy to overload yourself and for your enthusiasm to turn into a tendency to be overzealous. Achieving a balance may also require that you apply these rules with an appropriate pinch of salt. For example, you may feel branding (Rule 2) is superfluous, recording all talks via social media (Rule 7) may not suit everyone, and you may want to encourage unexpected material (Rule 8) by not telling speakers what to say (Rule 4).

However—and this is really the aim of all the rules—in order to make those decisions, you need to be reasonably well informed about what workshop organization will and could involve. The requirements and possibilities aren't always obvious and are broader than these ten simple rules (see [1], [3], [7], and [9]). Even if you go for the "simple" option, it is worth being aware of what opportunities or problems you are excluding or encouraging; for instance, if you were to ignore the benefits of social engineering (Rule 3), seating (Rule 6), or communications (Rule 5) or by being unprepared for timetable changes (Rule 10) and the workshop closing (Rule 9). Good luck! I have never got everything right at the same time!

## Acknowledgments

Thank you to co-organizers, funders (British Ecological Society, Microsoft Research, Microsoft Research Connections, Natural Environment Research Council [NERC]), participants, and organizations who supported workshops I've worked on.

## References

1. Corpas M, Gehlenborg N, Janga SC, Bourne PE (2008) Ten Simple Rules for Organizing a Scientific Meeting. PLoS Comput Biol 4(6):1–3. e1000080. doi: 10.1371/journal.pcbi.1000080
2. Michaut M (2011) Ten Simple Rules for Getting Involved in Your Scientific Community. PLoS Comput Biol 7(10): e1002232. doi: 10.1371/journal.pcbi.1002232 PMID: 22046114
3. Pavelin K, Pundir S, Cham JA (2014) Ten Simple Rules for Running Interactive Workshops. PLoS Comput Biol 10(2): e1003485. doi: 10.1371/journal.pcbi.1003485 PMID: 24586135
4. Bourne PE, Barbour V (2011) Ten Simple Rules for Building and Maintaining a Scientific Reputation. PLoS Comput Biol 7(6): e1002108. doi: 10.1371/journal.pcbi.1002108 PMID: 21738465

5. de Ridder J, Bromberg Y, Michaut M, Satagopam VP, Corpas M, et al. (2013) The Young PI Buzz: Learning from the Organizers of the Junior Principal Investigator Meeting at ISMB-ECCB 2013. PLoS Comput Biol 9(11): e1003350. doi: 10.1371/journal.pcbi.1003350 PMID: 24244148
6. Bourne PE (2007) Ten Simple Rules for Making Good Oral Presentations. PLoS Comput Biol 3(4): e77. doi: 10.1371/journal.pcbi.0030077 PMID: 17500596
7. Martin JL (2014) Ten Simple Rules to Achieve Conference Speaker Gender Balance. PLoS Comput Biol 10(11): e1003903. doi: 10.1371/journal.pcbi.1003903 PMID: 25411977
8. Ekins S, Perlstein EO (2014) Ten Simple Rules of Live Tweeting at Scientific Conferences. PLoS Comput Biol 10(8): e1003789. doi: 10.1371/journal.pcbi.1003789 PMID: 25144683
9. Bateman A, Bourne PE (2009) Ten Simple Rules for Chairing a Scientific Session. PLoS Comput Biol 5(9): e1000517. doi: 10.1371/journal.pcbi.1000517 PMID: 19779547
10. Budd A, Dinkel H, Corpas M, Fuller JC, Rubinat L, et al. (2015) Ten Simple Rules for Organizing an Unconference. PLoS Comput Biol 11(1): e1003905. doi: 10.1371/journal.pcbi.1003905 PMID: 25633715

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for a Community Computational Challenge

**Iddo Friedberg[1,2]*, Mark N. Wass[3], Sean D. Mooney[4], Predrag Radivojac[5]**

**1** Department of Microbiology, Miami University, Oxford, Ohio, United States of America, **2** Department of Computer Science and Software Engineering, Miami University, Oxford, Ohio, United States of America, **3** Centre for Molecular Processing, School of Biosciences, University of Kent, Canterbury, Kent, United Kingdom, **4** The Buck Institute for Research on Aging, Novato, California, United States of America, **5** Department of Computer Science and Informatics, Indiana University, Bloomington, Indiana, United States of America

* i.friedberg@miamioh.edu



**Citation:** Friedberg I, Wass MN, Mooney SD, Radivojac P (2015) Ten Simple Rules for a Community Computational Challenge. PLoS Comput Biol 11(4): e1004150. doi:10.1371/journal.pcbi.1004150

**Published:** April 23, 2015



**Funding:** The CAFA meetings at ISMB 2011 and 2013 conferences were supported, in part, by the US National Institutes of Health (NIH) grants R13 HG006079-01A1 and R13 HG007807 01 awarded to PR; and Office of Science (Biological and Environmental Research), US Department of Energy (DOE BER), grant DE-SC0006807TDD awarded to IF. Individual projects were supported, in part, by NIH R01LM009722 (SDM) and NSF ABI Innovation grant 1146960 (IF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

In science, the relationship between methods and discovery is symbiotic. As we discover more, we are able to construct more precise and sensitive tools and methods that enable further discovery. With better lens crafting came microscopes, and with them the discovery of living cells. In the last 40 years, advances in molecular biology, statistics, and computer science have ushered in the field of bioinformatics and the genomic era.

Computational scientists enjoy developing new methods, and the community encourages them to do so. Indeed, the editorial guidelines for *PLOS Computational Biology* require manuscripts to apply novel methods. However, it is often confusing to know which method to choose: which method is best? And, in this context, what does "best" mean?

To help choose an appropriate method for a particular task, scientists often form community-based challenges for the unbiased evaluation of methods in a given field. These challenges help evaluate existing and novel methods, while helping to coalesce a community and leading to new ideas and collaborations.

In computational biology, the first of these challenges was arguably the Critical Assessment of protein Structure Prediction, or CASP [1], whose goal is to evaluate methods for predicting three-dimensional protein structure from amino acid sequence. The first CASP meeting was held in December of 1994, following a "prediction period" where members of the community were presented with protein amino acid sequences and asked to predict their three dimensional structures. The sequences that were chosen had recently been solved by X-ray crystallography but had not been not published or released until after the predictions from the community were made. Since the first CASP, we have seen many successful challenges, including Critical Assessment of Function Annotation (CAFA) for protein function prediction [2], Critical Assessment of Genome Interpretation (CAGI) (for genome interpretation) [3], Critical Assessment of Massive (originally "Microarray") Data Analysis (CAMDA) (for large-scale biological data) [4], BioCreative (for biomedical text mining) [5], the Assemblathon (for sequence assembly), and the NCI-DREAM Challenges (for various biomedical challenges), amongst others [6].

Computational challenges also help solve new problems. While the original CASP experiment was developed to evaluate existing methods applied to current problems, other communities often look at other areas for which there are no existing tools. These challenges have spread successfully to industry, and companies such as Innocentive [7] and X-Prize [8] offer large prizes for solving novel questions.

Because these challenges are, on one hand, an exercise in community collaboration, and on the other, a competition, organizing a challenge is littered with difficulties and pitfalls. Having served as organizers, predictors, and assessors within several existing communities, we present ten rules we believe should be observed when organizing a computational methods challenge:

## Rule 1: Start with an Interesting Problem and a Motivated Community

Organizers of community challenges should start with an active community studying an important, non-trivial problem, and a good number of published tools that solve this or a similar problem using different approaches. Ideally, the challenge should be based on real data and the problem itself should be compelling. It is best to organize the community challenge around a meeting, whether adding the challenge on to an already scheduled event, or establishing a meeting especially for the challenge. Advertising a challenge without having the people to build on may doom your effort before it starts. Ensure you have a critical mass of predictors who are interested before you decide to move forward.

## Rule 2: Make Sure You Have Organizers, Data Providers, and Assessors Available before You Begin

The logistics of a challenge should be handled by separate entities, ideally comprising different people, to minimize potential conflicts of interest. These entities are the *data providers*, who give the testing data on which the methods are to be tested; the *assessors*, who assess the performance of the methods; the *organizers*, who provide the logistic infrastructure for the challenge, and the *predictors* are those who perform the predictions and whose predictions are assessed. Finally, there is a *steering committee*, composed of members who are knowledgeable in the field, but have no stake in the challenge. The members of the steering committee should offer different perspectives to the organizers on everything from rules to logistics, and thus help to better guide the challenge. The organizers should hold regular meetings with the steering committee, report progress, and identify possible faults along the way, and the roles should be sufficiently separated to ensure integrity. Everyone involved should be aware that there is a lot of work to be done over an extended period of time, and that during "crunch periods" such as challenge assessments, the work can take 100% of the time of several people over a few weeks. Even during calmer times, prepare for an extra workload that includes advertising, organizing, writing rules, developing metrics, and developing supporting software and web sites.

## Rule 3: Develop Reasonable Rules, but Be Flexible in Their Application, Especially the First Time

Work with your community and steering committee to come up with reasonable rules for the challenge, but understand that to recognize scientific impact, unforeseen changes will be required, particularly during the first iteration. These rules should be jointly developed by the organizers and the assessors and should be shared with the community for feedback. Learn from the first challenge, and change as necessary in future iterations to adopt rules that fit the questions at hand and the community's ability to address those questions.

## Rule 4: Carefully Consider Your Assessment Metrics

Good, unbiased assessment of the methods is critical to ensuring a successful challenge. Assessors should develop and publish their metrics early, and community input should be collected and used to refine them. The software that the assessors will use to evaluate predictions should be freely available. If possible, recruit assessors who are known and respected by your community. For some challenges, assessment is obvious, and only a single metric is needed. For others, the assessment methods can be more subjective and therefore contentious. If appropriate to your challenge, develop several complementary assessment methods. Take care to keep your assessment metrics easily interpretable. Metrics that are too complex and hard to explain can defeat the purpose of a community challenge.

## Rule 5: Have a Publication Plan

Try to have a publication plan prior to the challenge. Having the backing of a journal willing to publish papers from your challenge may help draw more people to the challenge. Of course, you should ensure that any manuscript is properly peer-reviewed. In CAFA, we availed ourselves of the special supplement mechanism provided by some journals [9]. Typically the papers are a publication of some of the participating methods, and those are authored by the method developers. A *flagship paper*, which provides a broader view of the challenge and participating methods, and authored by all participants, should also be included. You may use several publishers and different journals.

## Rule 6: Encourage Novelty and Risk-Taking

When creating a challenge, and especially with an ongoing challenge, predictors may gravitate toward marginally improving upon what worked in the past, rather than taking risky innovations. It is up to the organizers to encourage risk-taking, as that is where innovations usually originate. The challenges should have some novel edge to them to encourage the concurrent development of significantly novel methods. Also, the time given between challenges should be long enough to allow for the development of new methods, as well as substantial improvements to existing methods. (Typically 2–3 years between major challenge events). Finally, avoid "penalizing" methods that are not as competitive. This can be done by allowing authors to withdraw from the challenge or opt not to have their results published, thus allowing them to improve their methods and avoid a penalty from having a poor score publicly associated with their method.

## Rule 7: Build, Maintain, and Expand Your Community

Holding regular meetings based on your challenge builds community, encourages collaborations, and generally helps your effort become sustainable. Note that organizing meetings can seriously impact your time. Therefore, make sure that you and your collaborators are up for the effort. For more information, see *Ten Simple Rules for Organizing a Scientific Meeting* [10]. Have a meaningful and well-maintained website as a go-to resource for members of your community. Advertise your effort by presenting it at conferences, and over social media. Finally, seek feedback from your community after each event. Seeking feedback will help you understand what you are doing well, and what you may be able to improve. To facilitate a large number of honest responses an anonymous survey is the best way to gain feedback: use tools such as SurveyMonkey or SurveyGizmo. Also, do not allow your challenge to become stale. It may be exciting to gradually but constantly innovate by digging deeper and addressing more challenging aspects of your core problem.

## Rule 8: Seek Funding

Conferences and the challenge effort itself will need funding to sustain growth and existence. You should treat the challenge just like any other research project and seek sustainable funding for it. Work to convince funding agencies and commercial supporters that your challenge is timely and important and could make tangible contributions. Letters of support from challenge participants are crucial here. If possible, you should turn your results into new science where you and your community will test new emergent hypotheses. To make the challenge more transparent and help in the scientific research related to your challenge, you should provide assessment and other relevant software the community can use. Also, urge your participants to release their software to the public. Having tangible, useful products resulting from your challenge will serve the community, as well as help attract funding.

## Rule 9: Give Scientific Credit to the Predictors

One temptation of organizers of conferences, community challenges, or community challenge manuscripts is to somehow give the organizers scientific credit for the work of the data providers and predictors. Predictors and data providers should be celebrated as authors on manuscripts, speakers, and future organizers. Creating an environment where organizers (and not participants) are celebrated will only result in less impactful results, lower participation, and the overall quality of the ongoing challenge may be weakened. At the same time it is important to avoid ranking labs; instead, rank approaches: in the end, it is the competitive drive that motivates many of the predictors in these experiments. While this is well understood by both the community and organizers, calling the challenge and/or treating it like a personal competition will have the likely outcome of stifling risk.

## Rule 10: Prepare for an Incredible Ride, and Have Fun

If your effort is successful, you will be looked to by a community made up of dozens of groups, each group seeking to establish that their method is successful in the challenge. Naturally, in such a competitive environment, tensions and disagreements will arise—over the rules, the metrics, the challenge data, and anything that can be changed. Be patient and understanding, and most importantly, be attentive to criticisms and to the possibility of change. At the same time, after establishing rules, data, and assessment metrics, stick to your guns for the duration of the challenge round, unless you discover a major mistake that could render the challenge meaningless or grossly unfair. You can always fix lesser problems in the next round. You cannot, and therefore should not, aim to please everybody. Be patient and remember that the predictors are working hard and have a lot at stake, which may frustrate them when the assessment is not all they expected. Always remember that the challenge you are organizing is intended to improve methods through a friendly competition, and that you are involved in a community that, when all is said and done, should be a collegial one. Remember to have fun!

## Conclusions

Overall, we believe that if you follow these guidelines, you will be well on your way to helping improve tools and methods through community driven challenges. Make the scientific goals of the challenge abundantly clear and do not try to game the system to profit from the challenge itself. It is hard work, and may initially be unrewarding. The end result, however, can be as rewarding as any in science.

## Acknowledgments

The authors would like to acknowledge all CAFA participants. We are especially grateful to Patricia C. Babbitt, Steven E. Brenner, Michal Linial, Christine Orengo, and Burkhard Rost for their service on the CAFA steering committee.

## References

1. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins 1995; 23(3): ii–iv. PMID: 8710822
2. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T. et al. A large-scale evaluation of computational protein function prediction. Nat Methods 2013; 10: 221–227. doi: 10.1038/nmeth.2340 PMID: 23353650
3. CAGI. http://genomeinterpretation.org
4. Johnson KF and Lin SM. Call to work together on microarray data analysis Nature 2001; 411:885. PMID: 11418826
5. Hirschman L, Yeh A, Blaschke C, Valencia A. Overview of BioCreAtIvE: critical assessment of information extraction for biology. BMC Bioinformatics 2005; 6 Suppl 1:S1. PMID: 16468169
6. Costello JC and Stolovitzky G. Seeking the wisdom of crowds through challenge-based competitions in biomedical research. Clin Pharmacol Ther 2013; 93(5):396–398. doi: 10.1038/clpt.2013.36 PMID: 23549146
7. Innocentive, Waltham, MA, USA. http://www.innocentive.com/
8. Xprize. Culver City, CA, USA. http://www.xprize.org/
9. The Automated Function Prediction SIG 2011 featuring the CAFA Challenge: Critical Assessment of Function Annotations. Radivojac P, Friedberg I, Eds. BMC Bioinformatics 2011; 14(sup 3): S1–S15.
10. Corpas M, Gehlenborg N, Janga SC, Bourne PE. Ten simple rules for organizing a scientific meeting. PLoS Comput Biol 2008; 4(6):e1000080. doi: 10.1371/journal.pcbi.1000080 PMID: 18584020

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Organizing an Unconference

**Aidan Budd[1]*, Holger Dinkel[1], Manuel Corpas[2], Jonathan C. Fuller[3], Laura Rubinat[1,4], Damien P. Devos[5], Pierre H. Khoueiry[6]*, Konrad U. Förstner[7], Fotis Georgatos[8], Francis Rowland[9], Malvika Sharan[10], Janos X. Binder[1,11], Tom Grace[12], Karyn Traphagen[13], Adam Gristwood[14], Natasha T. Wood[15]**

**1** Structural and Computational Biology (SCB) Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, **2** The Genome Analysis Centre (TGAC), Norwich Research Park, Norwich, United Kingdom, **3** Heidelberg Institute for Theoretical Studies (HITS) gGmbH, Heidelberg, Germany, **4** Molecular Microbial Ecology Group, Institute of Aquatic Ecology, Universitat de Girona, Girona, Spain, **5** Centro Andaluz de Biología del Desarrollo (CABD), Universidad Pablo de Olavide, Sevilla, Spain, **6** Genome Biology (GB) Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, **7** Core Unit Systems Medicine, University of Würzburg, Würzburg, Germany, **8** Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg, **9** European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom, **10** Institute for Molecular Infection Biology, Universität Würzburg, Würzburg, Germany, **11** Bioinformatics Core Facility, Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Walferdange, Luxembourg, **12** Typeface & Lettering Design, Heidelberg, Germany, **13** ScienceOnline, Durham, North Carolina, United States of America, **14** Office of Information and Public Affairs, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, **15** South African National Bioinformatics Institute (SANBI), South African Medical Research Council Bioinformatics Unit, University of the Western Cape, Cape Town, South Africa

* aidan.budd@embl.de (AB); khoueiry@embl.de (PHK)





**Citation:** Budd A, Dinkel H, Corpas M, Fuller JC, Rubinat L, Devos DP, et al. (2015) Ten Simple Rules for Organizing an Unconference. PLoS Comput Biol 11(1): e1003905. doi:10.1371/journal.pcbi.1003905

**Editor:** Philip E. Bourne, National Institutes of Health, United States of America

**Published:** January 29, 2015



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** All authors are, or have been, involved in organising unconference or unconference-like sessions. Some of these events require participants to pay a registration fee.

## Introduction

An academic conference is a traditional platform for researchers and professionals to network and learn about recent developments and trends in a particular academic field [1–4]. Typically, the organizing committees and sponsors decide the main theme and sub-topics of the conference and select the presenters based on peer-reviewed papers [5]. The selected speakers usually share their research with a large audience by means of presentations and posters. However, the most stimulating discussions generally take place over coffee breaks when attendees can interact with each other and discuss various topics, including their own research interests, in a more informal manner [1, 6, 7], while expanding their own professional networks. An emphasis on facilitating such informal/networking interactions is a central focus of "unconventional conferences"—or "unconferences."

While many people may not yet have taken part in an unconference, the concept has been around for more than two decades. Events with unconference formats, beginning as early as 1985, include Open Space Technology, Foo Camp, BarCamp, Birds of a Feather, EdCamp, ScienceOnline, and many others. The success of these events has made the unconference format increasingly popular and widely known [8–11].

Unlike traditional conferences, an unconference is a participant-oriented meeting where the attendees decide on the agenda, discussion topics, workshops, and, often, even the time and venues. The informal and flexible program allows participants to suggest topics of their own interest and choose sessions accordingly. The format provides an excellent opportunity for researchers from diverse disciplines to work collaboratively on topics of common interest. The overarching goal for most unconferences is to prioritize conversation over presentation. In other words, the content for a session does not come from a select number of individuals at the front of the room, but is generated by all the attendees within the room, and, as such, every participant has an important role.

Advantages of the unconference format include: a focus on topics that are relevant to the attendees (because they suggested them), an opportunity for teamwork development, flexibility of schedule, and an emphasis on contributions from every participant. The relationships built during an unconference often continue well past the event. The interactions can lead to productive collaborations, professional development opportunities, and a network of resources and are very effective at building a community amongst participants. The unconference format, therefore, gives participants experience in working together, and this can change how they think about their day-to-day work.

A range of articles offer tips and advice for organizing and delivering aspects of scientific conferences and meetings or observations on features of successful meetings [5, 12, 13], including several from the *PLOS Computational Biology* "Ten Simple Rules" collection [14–16]. While the rules presented in this article are of particular relevance to the organization of unconferences, several of these points are also useful and complementary guidelines for organizing other kinds of events.

## Rule 1: How to Decide Whether to Run an Event As an Unconference or As a Traditional Conference

While there is no magic formula, reflecting on aspects such as participant numbers, venue size, expectations of attendees, and your overall objectives can be invaluable in deciding whether to run an event as an unconference or traditional conference. Unconferences are well suited to promoting interactions and networking between attendees as they allow a more flexible agenda. Discussion topics are shaped and influenced by participants, with exchanges of knowledge from many to many. This works particularly well when discussion groups are relatively small, creating a flexible, creative, and conducive environment for exchanges. A traditional conference, on the other hand, can be better suited to larger audiences, and when the focus of the meeting is more towards formal learning and knowledge sharing rather than involvement and interactions amongst participants. However, our experiences show that including unconference sessions in such events can be another valuable way of getting people involved, making connections, getting creative, achieving goals together, and developing a valuable platform for interactive knowledge exchange. It should also be noted that some successful unconferences are relatively large (e.g., ScienceOnline Together has 500 participants).

## Rule 2: Choose the Right Format

Depending on the mission and the goals of the participants, unconferences can be organized in many different ways. One example of an informal meeting is known as "Birds of a Feather"—these are events that usually accompany a traditional conference, where participants organize themselves to discuss topics without any pre-planned agenda, similar to "bar camps," where the program is rewritten or overwritten on-the-fly by the participants using whiteboard schedule templates.

Other examples involving project-driven events include those mainly focused on technology topics and that involve software project development, such as "hackathons." During such events, small sub-teams gather to work together on developing/addressing particular parts of a software project.

A little more organization is needed to arrange a "curated unconference" where topics and structures are collected by potential participants prior to the event. A group of organizers, in a transparent and open procedure, then sort through these ideas to build a structure of large and/or small-group discussion.

By forming smaller groups of participants to discuss different topics amongst each group, a "world café" style discussion allows participants to tackle several topics in a limited amount of time. At certain time intervals, every participant moves to a different table to participate in a specific discussion. Finally, all discuss the outcome of the different discussions under the moderation of the organizer.

In a "fishbowl" discussion, chairs are arranged in concentric circles with four to five chairs in the innermost circle (called the fishbowl), which channels the discussion as only participants in the fishbowl discuss the topic while others listen; participants wanting to join the vocal discussion approach the fishbowl and (via a mediator) replace one of the current members of the bowl.

Presentation styles at an unconference commonly include time limits, as exemplified by the "Ignite" and "Pecha Kucha" formats in which each presenter only has a very limited presentation time slot and slides advance automatically after 15 or 20 seconds, respectively. Such a format ensures that the presentations are succinct and fast-paced.

## Rule 3: Have a Clear Mission for the Meeting

Having a clear and visible mission statement can be a very effective way of focusing ideas for the content and structure of the event. It can turn collective minds to the development of a shared common goal that reduces emphasis on the individual and instead creates an event reflective of what the group needs and wants. From our experience, there are two major reasons why people attend unconferences: (1) to interact with many people of shared interests and (2) to learn useful information or skills related to their activities (often focused on their own career progression). A clear mission is a useful way of focusing the expectations of participants to the goals of the meeting. It can help to create an environment conducive to valuable and appropriate learning, and can guide discussions beyond a mere brainstorming session. Decisions about the focus and content of specific sessions become less subjective and remain transparent when the decision criteria align with the overarching goal of the meeting.

## Rule 4: Minimize the Lecture-Style Presentations

One of the defining features of an unconference is its inversion of the common features of more traditional meetings, in particular academic conferences. A common aspect of traditional meetings is the formal presentation (i.e., lecture style) with communication directed from one, typically a senior and powerful member of the community, to many others who listen passively and do not have much opportunity to actively interact with the presenter's ideas. In contrast, unconferences typically minimize the use (and duration) of conventional presentations and prioritize cooperative knowledge. This means that the session content comes from the shared experiences and expertise of all participants in the room and not just from the front of the room. The idea that no individual person has all the answers promotes a spirit of generosity, interaction, and respect amongst all participants. Every voice is valued.

## Rule 5: Involve Participants in Planning and Structuring of the Event

Participant-centric thinking is perhaps the key feature that differentiates unconferences from more traditional meetings. Empowered participants, who know that they can directly influence and contribute to the structure and content of a meeting, tend to be much more invested in its success and outcome. However, the events still involve a certain amount of planning and infrastructure [14] and paying attention to details such as required equipment, venue, network connectivity, power outlets, and catering can have a large impact on the success of the event. Managing the flexibility of an unconference with appropriate logistical organization can avoid wasting time and, thus, avoid frustration for both the participants and organizers.

Participation is also where much of the enthusiasm and excitement of such meetings comes from, and there are many ways in which contributions can be facilitated. If a core group of organizers takes the lead in planning the event—including the program—then participants can focus on taking part in the discussion of ideas for sessions, content, or form of the unconference (see "Rule 2" for a variety of discussion formats and styles) instead of dealing with frustrating details. To ensure that the logistical arrangements are carried out prior to the event, the role of each organizer should be clearly communicated. As such, it may be beneficial to appoint one individual who coordinates the activities and is responsible for following-up on important preparations. Furthermore, the agenda should be visible to all participants before the unconference takes place and should include essential information such as the theme, sub-topics, time allowance, and contact information. These standard preparations allow the participants to arrive well informed and also create an opportunity for each participant to decide on how they may want to contribute to the unconference.

During the wrap-up of the event, any suggestions and feedback regarding the overall unconference events can be discussed and the theme of the next unconference can be decided. The goals of the next event will guide the planning and participants will be able to volunteer to be part of the new group of organizers. Finally, encouraging facilitators to include people who they know have interesting contributions to make ensures a core of contributors and promotes a lively discussion.

## Rule 6: Provide an Open, Relaxed Atmosphere

In order to make an unconference a success, the atmosphere of the event should be relaxed, open, friendly, and fun. This will ensure that all participants, especially those joining for the first time, feel welcome and respected. Creating and encouraging a casual and relaxed environment is favourable for everyone involved because it facilitates interaction and communication. To promote a relaxed atmosphere, think carefully about the layout of the venue. This includes the size of the room and the placement of tables and chairs; for example, arranging tables for small group discussions or placing chairs in a semi-circle or U-shape for group discussions. A good set-up not only fosters discussion but also has a positive impact on the overall quality of the unconference by strengthening the personal experience.

The organizers, as well as participants who have attended previous unconferences, should reach out and welcome newcomers to the format. By modelling conduct and values through their interactions with other participants both before and during the event (particularly at the start), they can strongly influence the way in which people interact with each other.

An effective way to encourage communication and participation is through ice-breaker activities during the early stages of the event. Small group activities are especially helpful since many participants may initially find it easier to interact actively in smaller, more intimate groups. This also helps new attendees meet new people and start to build relationships in a casual manner.

Fear of public speaking, questioning, and debating are common in all academic fields and communities. Unconferences aim to overcome these fears by creating an environment of respect that helps all participants gain self-confidence. Nominating capable, guiding facilitators who are able to ensure respectful communication throughout the meeting can achieve this goal. The facilitators should encourage all participants to share their own thoughts, listen to others' comments, and—most importantly—consider all contributions. Repeating the name of a participant linked to a developed idea gives this participant a boost in self-confidence. However, in some cases, it may also mean that over-confident participants need to be "moderated" to provide enough time and space for the least confident participants to contribute voluntarily. Therefore, while diverse opinions are welcomed (and often result in stimulating discussions), the focus at an unconference is on how these different opinions are communicated. Good facilitators will create a natural atmosphere of mutual respect and trust.

## Rule 7: Trust Your Community

Unconferences prioritize focusing on, and engaging with, everyone who chooses to get involved in the event. This is in contrast to more traditional meetings, where the focus is much more on what the organizers have planned and the scheduled session presenters. Thus, in an unconference format, responsibility for the success of the event is more equally distributed across all participants. This shift of responsibility away from the organizers can initially seem intimidating, as it might seem like there are fewer ways to influence the success of the event. The lack of control can be difficult to accept, particularly for those who tend to micromanage. In an unconference format, the organizers will be successful if they trust the community to work with them to make the event a success. This power shift is worth embracing, rather than resisting, as it brings many exciting and energizing opportunities. Sharing leadership with the participants will create an atmosphere of personal empowerment, individual responsibility, and group ownership of the events.

This is perhaps not surprising; almost everyone choosing to participate in an unconference does so to personally benefit from the event. When given the chance to influence the success of the event, the attendees count this as a benefit in addition to the content of the unconference itself.

Another benefit is that the workload of an organizer may be reduced if it can be shared amongst a group of volunteers. Finally, trusting in the community makes it easier and less risky to experiment with novel formats and topics. Even when these experiments do not work out as planned, the very act of trying new ideas by involving, engaging, and trusting in participants brings the community closer together and delivers its own kind of success in terms of networking and community building. Learning to trust the community is key to embracing and enjoying the special character of these events.

## Rule 8: Communication Is Key to Your Event; Make it As Easy, Unambiguous, and Transparent As Possible

Engaging in communication is one of the reasons why people choose to come together for any meeting. One main characteristic of unconferences is the emphasis on interactive communication that gives all participants a chance to have their contributions heard by others. To this end, make use of multiple existing collaborative tools that assist in the communication before, during, and after an event. For example, a wiki can be very helpful in giving participants the chance to get involved in the organization of the event in advance—including idea and topic collections, scheduling sessions, taking care of the infrastructure of an event, as well as finding accommodation and ride shares for low-cost events.

Several tools exist to help with jotting down notes or minutes during a session: classic white boards and colored pens can be useful to collect suggestions and develop ideas together; even getting participants to scribble their thoughts down on paper tablecloths (which is a low-cost and low-tech collaborative tool with great haptic feedback) has proven to be handy and fun. The final work can be photographed and the pictures made available online later. Web-based collaborative real-time editors like Etherpad (http://etherpad.org/) can be helpful to conceptualize thoughts and to track discussions, as they can be edited by multiple people in parallel and can be used afterwards as an equivalent to conference proceedings. However, these Web-based editors require a working Internet connection throughout the event, which may not be practical at each event. Social media such as Twitter can also be utilized to share topics, progress, statements, or questions with people who are not present at the session. Here it is important to agree on a short, but distinctive, hashtag as soon as possible to enable people to follow and keep track of the tweets. A Tweetwall—a large screen or a projector displaying the most current tweets associated with the event's hashtag—can also be entertaining and informative.

## Rule 9: The Journey Is As Important As Its Destination

A great way to extract the collective expertise, knowledge, and experience of attendees during unconference sessions is to encourage participants to identify and work together towards a common goal, and to document how they attempted to get there. Any given event will rarely provide the time needed to take a goal or project from beginning to end; however, we have seen unconferences serve as excellent ways of brainstorming, developing initial plans, creating the outline for a project, and gathering together a group of enthusiastic collaborators.

It is important to have tools that allow attendees to share the resources, ideas, and challenges of the session conversations. Documenting content can be an effective way to engage people and also to further the legacy of the unconference session beyond the confines of the room. Such an approach provides a way for participants to reflect on the collective learning and thinking that took place, as well as providing the means to evaluate the success of the discussion. It is unlikely within the time constraints of a session or single event that participants will come up with "the one final answer" to a particular problem or challenge. Therefore, providing a collaborative tool to record the development of ideas during the unconference session is important. The documentation of the session is a resource for reflecting on the work done, enabling participants to think about the issue in different ways, allowing others to see the progress of the discussion, establishing ideas for future events, and building a network of collaborators. In other words, the recording of the journey yields many benefits, even if you do not reach your final destination.

## Rule 10: No Idea Is Too Trivial

When a diverse group works together, some individuals will be good at big picture suggestions and others will emphasize details. Both are needed and both should be encouraged. While discussions of new ideas often begin at the conceptual level, contributions that may seem trivial or detail-oriented in the moment can also be important to a project's ultimate success. Thus, to avoid missing out on important contributions, it is essential to include even the seemingly trivial remarks or ideas. A good way to do this is to write down all ideas and suggestions, so that later they can be sorted and considered. Do not rule out anything when it is first suggested because brainstorming becomes the most productive when any idea that comes to mind is communicated without prior judgment of its value. One person's unusual idea may spark the way forward.

## Final Thoughts

There is not one "right" way to organize an unconference, but there are certainly things to be sure to include (and to avoid!) so that the event is as successful as possible. Perhaps the key is thinking of the event as "we" instead of "me."

## Crowdsourcing the Writing of This Article

The authors wanted to base the opinions and advice provided in this article on experience of diverse unconferences. By doing this, rather than relying on the opinions of a small group of authors, we hoped that the content would be useful to a wider range of people. Thus, we crowdsourced the content by contacting organizers of a range of unconferences and similar events and inviting them to join us as authors. We also invited as authors all participants of a Birds of a Feather session focused unconference at the ISMB/ECCB 2013 meeting in Berlin, including also those who contributed to this session remotely via Twitter. Finally, we also invited all organizers of the Heidelberg Unseminars in Bioinformatics series of events [17] to join as authors, as several of the initiators of this article are members of that group.

We began the crowdsourcing by collecting a list of possible rules for the article via a git-controlled repository [18]. This list was then trimmed to reduce redundancy and overlap, and all authors voted to identify the initial set of ten rules to be included in the article. Small teams of authors collaborated to write content for each rule using a Piratenpad (https://www.piratenpad.de/), an online collaborative writing tool similar to an Etherpad. Native English speakers amongst the authors then processed this first draft to provide a common tone and language to the article. The resulting draft was then discussed by all authors, distributed as a Word document, and edits were implemented on the basis of this discussion by one of the authors until a consensus version of the text was agreed upon and submitted to the journal.

Authors are listed in the byline in the order in which they made edits to the manuscript.

## References


1. Cutting WA (1995) How to do it. Participate in an international conference. BMJ 310: 249–251. doi: 10.1136/bmj.310.6974.249
2. Petsko GA (2006) The highs and lows of scientific conferences. Nat Rev Mol Cell Biol 7: 231–234. doi: 10.1038/nrm1832 PMID: 16421519
3. Ramdayal K, Stobbe MD, Mishra T, Michaut M (2014) Building the future of bioinformatics through student-facilitated conferencing. PLoS Comput Biol 10: e1003458. doi: 10.1371/journal.pcbi.1003458 PMID: 24499938
4. Alberts B (2013) Designing scientific meetings. Science 339: 737. doi: 10.1126/science.1236324 PMID: 23413322
5. Potvin JH (1983) Planning and organizing an annual conference. IEEE Trans Prof Comm PC-26: 123–152. doi: 10.1109/TPC.1983.6448156
6. McIntyre E, Millar S, Thomas F (2007) Convening a conference—facilitating networking among delegates. Aust Fam Physician 36: 659–660. PMID: 17676193
7. Pierce G (2014) The dilemma of attending (or not) scientific conferences. Can J Physiol Pharmacol 92: v. doi: 10.1139/cjpp-2013-0412
8. Boule M (2011) Mob Rule Learning: Camps, Unconferences, and Trashing the Talking Head. Medford, New Jersey: Information Today Inc. 224 p.
9. Owen H (1998) Expanding Our Now: The Story of Open Space Technology. San Francisco: Berrett-Koehler Publishers. 147 p.
10. Owen H (2008) Open Space Technology: A User's Guide (Third Edition). San Francisco: Berrett-Koehler Publishers. 192 p.
11. (2006) Foo's paradise. Nature 442: 848. doi: 10.1038/442848a PMID: 16929260
12. Wyatt J (1999) Organising a medical conference. J Accid Emerg Med 16: 223–226. doi: 10.1136/emj.16.3.223 PMID: 10353055
13. (2012) A day off in Denmark. Nat Rev Microbiol 10: 667. doi: 10.1038/nrmicro2892 PMID: 23136691
14. Corpas M, Gehlenborg N, Janga SC, Bourne PE (2008) Ten simple rules for organizing a scientific meeting. PLoS Comput Biol 4: e1000080. doi: 10.1371/journal.pcbi.1000080 PMID: 18584020

15. Gichora NN, Fatumo SA, Ngara MV, Chelbat N, Ramdayal K, et al. (2010) Ten simple rules for organizing a virtual conference —anywhere. PLoS Comput Biol 6: e1000650. doi: 10.1371/journal.pcbi.1000650 PMID: 20195548

16. Pavelin K, Pundir S, Cham JA (2014) Ten simple rules for running interactive workshops. PLoS Comput Biol 10: e1003485. doi: 10.1371/journal.pcbi.1003485 PMID: 24586135

17. Fuller JC, Khoueiry P, Dinkel H, Forslund K, Stamatakis A, et al. (2013) Biggest challenges in bioinformatics. EMBO Rep 14: 302–304. doi: 10.1038/embor.2013.34 PMID: 23492829

18. Ram K (2013) Git can facilitate greater reproducibility and increased transparency in science. Source Code Biol Med 8: 7. doi: 10.1186/1751-0473-8-7 PMID: 23448176

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules to Achieve Conference Speaker Gender Balance

**Jennifer L. Martin***

Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia

Recently, the quantum molecular science world was in uproar [1,2]. The preliminary list of approximately 25 speakers for the International Congress of Quantum Chemistry (ICQC) was published online, with no women speakers listed. One reaction to this list was to set up a petition to "condemn gender-biased discriminatory practices of which ICQC-2015 is the most recent example" [3]. This resulted in an apology and a new speaker list with six women speakers [4].

Sadly though, this is not an isolated incident: men-only invited conference speaker lists are all too common [5].

How can we get gender balance right? To begin with, it's worth reminding ourselves why gender balance is important.

First, it's critical for the future of science that young women and men can see real evidence that scientists can succeed regardless of gender. So, if we are going to encourage women into careers in science we need also to provide role models for them to aspire to. We need to show that being a woman and being a successful scientist are not mutually exclusive. One way of doing that is to give women scientists a platform to present their research. If we don't address gender balance in speaker programs, we will continue to normalise a gendered stereotype of scientific leadership. Then when crunch time comes, women will continue to leave in far greater numbers than men [6–9] in part because they see no path ahead for themselves. And that means scientific research potentially loses half of its brightest talent.

Moreover, a speaking invitation contributes enormously to the profile of a researcher. By extending more invitations to women and other under-represented sections of the academic community, we provide a boost to their visibility and their track record. This will help them to progress by raising their national and international profile and help support their applications for grants, academic positions, and fellowships.

Finally, conferences and symposia are great ways of generating new collaborations, new ideas, and new directions in science. If we keep inviting the same people, and the same types of people, over and over again, we limit the diversity of thought and, potentially, the opportunities for innovation.

So, here are ten simple rules to achieve conference speaker gender balance.

## Rule 1: Collect the Data

Count the number of women and men attending a conference, or the number of women and men who have membership of a professional society, or the number of women and men who are employed or studying at a University department. If the same conference/seminar series has been running for a number of years, averaged data could be used (over the past five years, for example). When running a conference for the first time or collecting information about society membership, make sure to include gender as one of the questions to allow this base rate to be generated. Use the information to determine the gender balance of the conference, seminar series, or department. Of course, this may change over time, so it's worth checking every few years.

## Rule 2: Develop a Speaker Policy

A speaker policy captures what the committee is trying to achieve for its members and audience when putting together the speaker program. It can also help the committee measure outcome. A policy may state, for example, that the conference committee wants to achieve a gender balance of speakers that roughly reflects that of its audience. Depending on the conference or meeting, the policy might include scientific diversity, geographical distribution, ethnicity, and level of seniority in the speaker policy. If you are not sure what a conference policy looks like, check examples written by others, such as the Lorne Proteins conference [10] or the Crystal29 conference [11]. The policy can be quite simple and, yet, still effective. Data from Rule 1 above will feed into and perhaps help modify the policy, but a policy should be developed immediately. An anti-harassment statement should also be included in the conference policy [12].

## Rule 3: Make the Policy Visible

It's no use having a policy if no one knows about it. Make it visible. Put it online for everyone to see. Make a direct link to it on the conference or symposia website and put it on your Facebook page. Provide it to the organising committee, the program committee, the society executive, and the departmental research committee. Send it to the chairs of the sessions, send it to the invited speakers. Make sure everyone knows right from the start that the conference committee is serious about getting gender balance right. Don't make gender balance an afterthought.

## Rule 4: Establish a Balanced and Informed Program Committee

If the conference program committee is not diverse, then neither will be the speaker

**Citation:** Martin JL (2014) Ten Simple Rules to Achieve Conference Speaker Gender Balance. PLoS Comput Biol 10(11): e1003903. doi:10.1371/journal.pcbi.1003903

**Published** November 20, 2014



**Editor:** Philip E. Bourne, National Institutes of Health, United States of America

**Funding:** I am funded by the Australian Research Council through an Australian Laureate Fellowship (FL0992138), and I am also an Australian National Health and Medical Research Council honorary Fellow (455829). The funders had no role in the preparation of the manuscript.

**Competing Interests:** This article is about achieving gender balance at conferences, because women are often underrepresented in invited speaker lists. I am a woman, and so could be considered to have a conflict of interest.

* Email: j.martin@imb.uq.edu.au

list. When I've asked male members of conference organising committees, or men in the audience, about poor gender balance or even good gender balance, they invariably tell me they don't notice the gender balance one way or the other. So to avoid the potential issue of gender blindness, make sure that those inviting and selecting the speakers (the program committee or symposia chairs) are familiar with the conference policy and that the program committee itself is diverse, informed, and gender balanced [13].

## Rule 5: Report the Data

The next step is to see how well the conference, speaker series, or symposium meets its stated policy goals. To do this, calculate the percentage of women in the list of invited speakers. And do the same for the selected speakers. How do these numbers compare with the percentage of women in the audience? If the percentage of women speakers is consistently lower than that of the audience, then maybe it's time to overhaul the policy. Maybe the gender balance of the program committee needs to be changed. In any case, report the numbers on the website, on the same page as the policy. Comment on the data. For example, "Our stated policy is to achieve a gender balance of >40% women in our speaker list (50% of our delegates are women). Overall, 35% of our speakers were women. This is good, but we can do better." Ask for feedback.

To go one step further, you could establish a website to "crowdsource and collate the gender breakdown of conferences" [14] to identify and promote conferences that best support gender balance.

## Rule 6: Build and Use Databases

Some people find it difficult to come up with names of women speakers, compared with men speakers. Some say there aren't enough senior or mid-career women in the field to get a balanced program. When I got this response last year after querying a proposed speaker list, I arranged to meet with the organiser to brainstorm a new list. We collected enough names in one afternoon to fill two or three conferences. The list of potential women could include younger, up-and-coming women who would benefit from the exposure.

It's also worthwhile looking through lists of women scientists that have been compiled to help conference organisers—see, for example [15–17]. This is by no means an exhaustive list of sites, so please add more. And if you can't find a list in your field, consider compiling one yourself.

## Rule 7: Respond to Resistance

Expect to meet resistance. Most criticisms are easily addressed by establishing a dialogue with those who are critical about establishing a policy, and you can prepare in advance.

Some will say the most important thing is not diversity or the number of women speakers, the most important thing is having a high-quality program. "We only select the best speakers." Addressing gender balance is not inconsistent with a high-quality program. Perhaps point them to the implicit association site [18].

Similarly, some will say the most important thing is diversity of thought, not speaker diversity. Diversity in life experience equals diversity of thought. Again, having a gender-balanced program is not inconsistent with diversity of thought. On the flip side, inviting the same people over and over again does not address diversity of thought.

Some will say, yes let's have a policy, but let's not make it public because that makes it look like we've had a problem in the past and are now apologising for it. There is no point having a policy if no one knows about it. Put it online. See Rule 3.

Some will say that a policy isn't needed because gender balance is achieved already. Check the data. See Rule 1. Maybe gender balance is OK, but it's important to ensure that invisible inequities do not prevail.

## Rule 8: Support Women at Meetings

Women often have primary caring responsibility for children. This can limit their ability to travel and to attend conferences. Professor Jonathan Eisen (UCD) has stated: "If you're going to spend money on an open bar instead of childcare…you should rethink what you're doing" [19]. Some universities are now offering travel support for partners or nannies to attendees who would otherwise not be able to accept conference speaking invitations. Perhaps conferences could do the same. Ideas on why women don't accept invitations and how to support their attendance, such as providing a childcare center and avoiding gendered language, have been outlined [20].

## Rule 9: Be Family-Friendly

In those cases where the conference is large enough, and the number of attendees bringing children is significant, it may be possible to provide a family room. This allows delegates with children to watch conference presentations via video link. Also consider carefully the social events to be scheduled at your conference. Make sure these are appropriate.

## Rule 10: Take the Pledge

Finally, the most important and powerful step of all. When you are invited to help organise, attend, or speak at a conference, ask to see the conference speaker policy before you accept. If there isn't one, which is usually the case, offer to help draft one. You could also ask to see the list of invited speakers and if there isn't a reasonable gender balance, just say no. That's what a group of Scandinavian men have pledged: to say no thanks, when there are no/few women speakers [21].

You could also sign the online petition set up by Virginia Valian and Dan Sperber [22] in which "signatories commit to accepting talk invitations only from conferences that have made good-faith efforts to include women."

So, those are the ten simple rules.

One day, hopefully not too far away, I'd like to think we won't actually need conference speaker policies anymore. The process of selecting and supporting a broad, diverse, balanced list of high-quality speakers will be as automatic as flicking to the next slide on a PowerPoint presentation.

## Acknowledgments

A big thank you to all the tweeps who suggested I modify my blog post into this list of ten simple rules, after yet another conference speaker program with no women listed.

Twitter: @JennyMartin_UQ

Blog: cubistcrystal.wordpress.com/

Portions of this article were previously published at cubistcrystal.wordpress.com/show-me-the-policy.

## References


1. Gibney E (17 Feb 2014) Chemists call for boycott over all male speaker line-up. Nature News Blog. Available: http://blogs.nature.com/news/2014/02/chemists-call-for-boycott-over-all-male-speaker-line-up.html. Accessed 20 June 2014.
2. Editorial (2014) Rethink your gender attitudes. Nat Mater 13: 427.
3. Krylov A (2014) Stop gender discrimination in science. Petition closed 18 March 2014. Available: http://www.change.org/p/scientific-community-stop-gender-discrimination-in-science. Accessed 20 June 2014.
4. Women in theoretical chemistry, material science and biochemistry (2014) The community protests gender discrimination at professional meetings. Available: Accessed 20 June 2014.
5. Eisen J (2014) Diversity in STEM. The Tree of Life. Available: Accessed 20 June 2014.
6. Handelsman J, Cantor E, Carnes M, Denton D, Fine E, et al. (2005) Careers in science. More women in science. Science 309: 1190–1191.
7. United States National Academy of Sciences (2007) Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering. Washington DC: National Academies.
8. Pain E (11 Feb 2014) More action needed to retain women in science. Available: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2014_02_11/caredit.a1400036. Accessed 20 June 2014.
9. Marchant T, Wallace M (2013) Sixteen years of change for Australian female academics: progress or segmentation? Australian Universities Review 55: 60–71.
10. Lorne Conference on Protein Structure and Function Organising Committee (2014) Conference policies. Available: http://www.lorneproteins.org/policies/. Accessed 20 June 2014.
11. Crystal29 Conference Organising Committee (2014) Conference policy. Available: http://www.crystal29.com/conference-policy/. Accessed 20 June 2014.
12. The Ada Initiative (2014) Anti-harassment work, including conference anti-harassment policy. Available: https://adainitiative.org/what-we-do/conference-policies/. Accessed 20 June 2014.
13. Casadevalla A, Handelsman J (2014) The presence of female convenors correlates with a higher proportion of female speakers at scientific symposia. mBio 5: e00846–13.
14. Holt K (2014) Gender balance in scientific conferences (Australia). Available: https://sites.google.com/site/aussiescience/. Accessed 20 June 2014.
15. American Physical Society (2014) Women speakers list. Available: http://www.aps.org/programs/women/speakers/. Accessed 20 June 2014.
16. American Society of Cell Biology Women in Cell Biology Committee. Speaker referral service. Available: http://www.ascb.org/index.php?option=com_content&view=article&id=94&Itemid=147. Accessed 20 June 2014.
17. EMBO Women in Life Science (2013) WILS Database of Women in Science. Available: http://www.embo.org/science-policy/women-in-science/wils-database-of-women-in-life-sciences. Accessed 20 June 2014.
18. Project Implicit (2011) Project Implicit Social Attitudes. Available: http://implicit.harvard.edu/. Accessed 20 June 2014.
19. STEM Women Interview with Jonathan Eisen (26 Feb 2014) STEM women: How men can help. STEM Women. Available: http://www.stemwomen.net/jonathan-eisen/. Accessed 20 June 2014.
20. King T (23 Sept 2013) Women Speakers at Events. TKingDoll Revolution. Available: http://www.tkingdoll.com/2013/09/women-speakers-at-events.html. Accessed 20 June 2014.
21. Tackanej. Men say no, thanks. #TackaNej. Available: http://tackanej.se/men-say-no-thanks/. Accessed 20 June 2014.
22. For Gender Equity Team (25 Sept, 2012) Commitment to gender equity at scholarly conferences. Available: http://www.gopetition.com/petitions/commitment-to-gender-equity-at-scholarly-conferences.html. Accessed 20 June 2014.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Running Interactive Workshops

**Katrina Pavelin, Sangya Pundir, Jennifer A. Cham***

European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

## Introduction

Do you have a difficult problem to solve? Are you writing a grant proposal involving several stakeholders? Do you want to gather user feedback on a resource, tool, or service? Or perhaps you need to improve a process or way of working in a team? To address problems such as these, we recommend holding an interactive workshop. We find that the dynamic nature of such workshops encourages creative thought and can quickly yield ideas and solutions [1].

Our ten simple rules aim to empower you to design and lead your own successful interactive workshops. We define an "interactive workshop" as a structured set of facilitated activities for groups of participants who work *together* to explore a problem and its solutions, over a specific period of time, in one location. Participants can be users, potential users, team members, customers, or stakeholders. Figure 1 shows a typical layout; note if you have five or fewer participants, consider having just one group.

An interactive workshop is distinct from a standard meeting because it aims to stimulate creativity through collaborative working. A meeting, in contrast, usually involves planning and reporting work with attendees sharing their individual points; it does not involve group activities "live" in the meeting. Also, meetings may be an hour or less, whereas the minimum time needed for an interactive workshop is 2–3 hours.

We also do not include training workshops, talks, or seminars in our definition, because from our experience they tend to comprise lectures and tutorials, not brainstorming activities. However, it could be possible to have an interactive workshop session as part of a longer training course.

## Rule 1: Decide Whether an Interactive Workshop Is the Right Choice

Interactive workshops can be useful in many situations, with both internal and external participants; see [2] for an example case study. Interactive workshops may be suitable for:

- gathering ideas for research grant proposals;
- ascertaining user requirements for bioinformatics services;
- generating ideas for designing web/software interfaces;
- solving problems, such as process improvement or work strategies;
- deciding priorities, strategy, and vision;
- improving working relationships through team building, such as part of retreats.

Before you start planning, it is important to determine whether an interactive workshop is the right choice. For instance, they are not usually advisable at the inception of a project when you need to identify the goals. Organisation leadership, policy, and many other factors may determine this. However, once objectives have been identified and agreed upon, then an interactive workshop can be valuable to explore *how* to meet them.

Interactive workshops are also unsuitable when you have firm alternatives to evaluate (like mock-ups for a website). It would be better to get individual feedback and then collate the results.

Another consideration is that interactive workshops can require extra time and resources to plan and deliver because activities, templates, and materials need to be prepared in advance and more people may be required for facilitating the activities. We suggest if there are significant constraints, especially short timescales, it may be more appropriate to hold a standard-format meeting [3].

## Rule 2: Choose Participants Carefully

Descriptions of your target groups (or "user profiles" [4]) may help guide your choice of participants for an interactive workshop. Aim for diversity in experience, opinions, seniority, and interests. For external participants, send out an electronic "screener" survey to find out if they truly represent your target profiles and use this information to assign groups for the activities. You may wish to split colleagues into different groups because separating people who usually work together exposes them to alternative perspectives and new thinking, thus stimulating creativity.

If you are not able to choose the participants yourself, you will still need to find out who they are and see how their perspective and/or background fits with the objective(s) of the workshop. Consequently, you may need to factor this into your analysis of the outcomes.

## Rule 3: Identify Suitable Activities

Before you start planning, think about how you (as the person managing the delivery of the interactive workshop) will present the outcomes in a talk or report. Consider: What are the tangible aims of the workshop? What specific information do you need to capture? Tailor the activities to these specific goals. Where possible, use engaging activities, such as "game-storming" techniques [5], to motivate your participants. By using visual metaphors and games you may further encourage creative thought, as this allows the rules of everyday life to be suspended

**Citation:** Pavelin K, Pundir S, Cham JA (2014) Ten Simple Rules for Running Interactive Workshops. PLoS Comput Biol 10(2): e1003485. doi:10.1371/journal.pcbi.1003485

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** February 27, 2014



**Funding:** KP and JAC's work was supported by the European Molecular Biology Laboratory (EMBL). SP's work was supported by the UniProt project, funded by National Institutes of Health (NIH) grant 1U41HG006104-01 and additional support from EMBL and the NIH Gene Ontology (GO) grant 2P41HG02273-07. The funders had no role in the preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: jennifer.cham@ebi.ac.uk

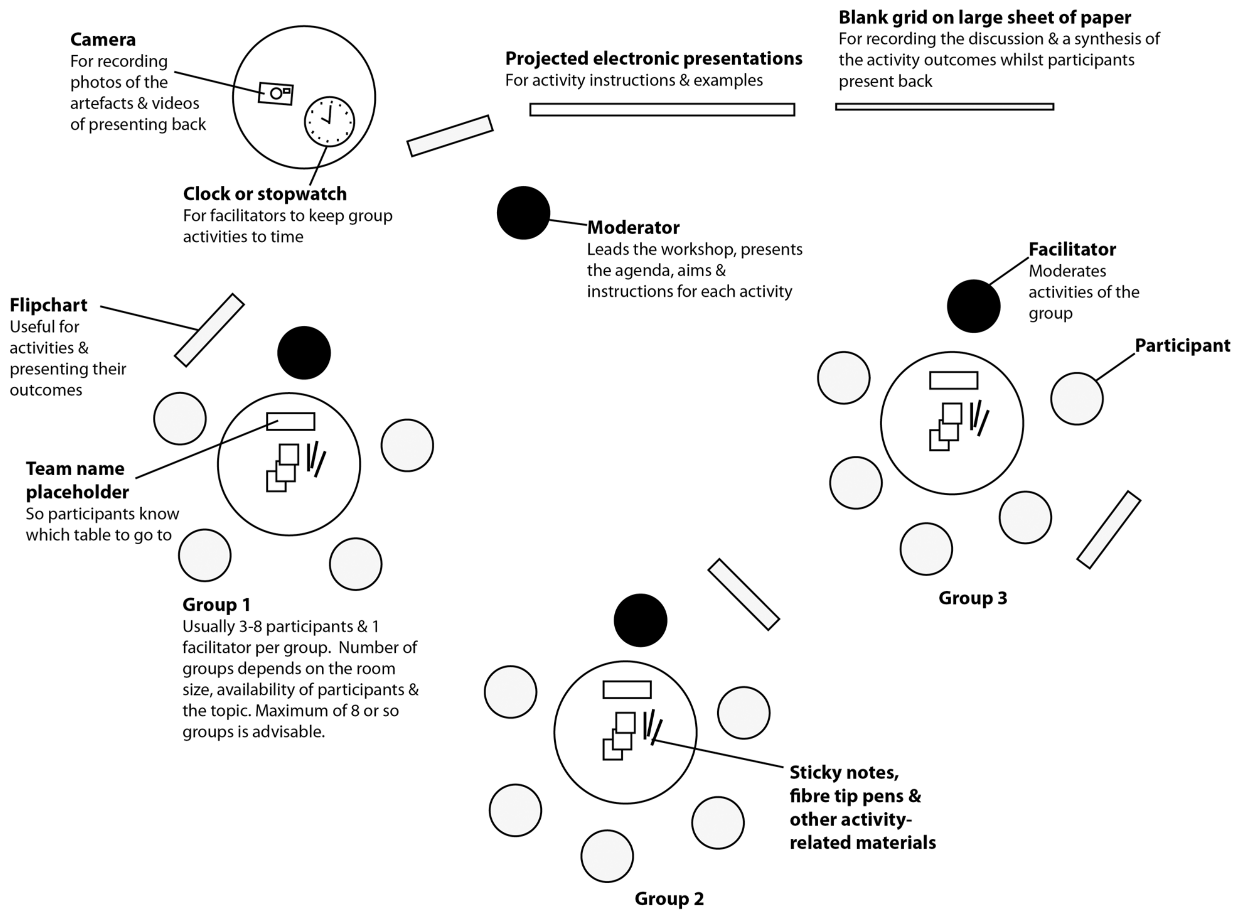



**Figure 1. Example room layout for an interactive workshop.** This bird's-eye view shows the setup for supporting group-based, facilitated activities around a specific topic, problem, or project. The moderator oversees the workshop with the help of facilitators, who are briefed in the aims and methods of the activities. Alternatives include "circuit training" layout, where each table is an activity station and the participants move around the room.
doi:10.1371/journal.pcbi.1003485.g001

for exploring problems in new, sometimes unorthodox, ways.

Some participants may be sceptical about the value of "just playing games." However, once you clearly explain the aim of each activity and how it is tailored to solve the problem, most participants will engage positively. For a checklist of how to introduce an activity to a group, see [6].

## Rule 4: Identify Facilitators and Brief Them

Facilitators coordinate and assist group discussions and activities during an interactive workshop. You will need to carefully brief these helpers to ensure they know how to moderate and are familiar with the aims and practicalities of the activities that are scheduled. You may also wish to have an overseeing moderator who presents the aims, agenda, and activity instructions (Figure 1).

Most importantly, you should emphasise that facilitators need to be impartial coordinators: neither contributing ideas, nor evaluating them, but rather encouraging input from participants in their group. They keep discussions on time, and remind participants to note down all their points, sometimes actually doing this for them whilst they are speaking. Facilitators may also note further ideas onto the artefacts as they arise during the presenting-back phase and subsequent discussion with everyone in the room.

It is not essential for facilitators to have substantial domain expertise, but it may be helpful for them to have at least a basic understanding of the concepts being explored. For example, for activities to prioritise items it helps if facilitators can clarify what the items are if participants have questions, or give specific examples as illustrations if needed.

On the day of the workshop, the facilitators (or moderator) will monitor the groups and, if there is time, they may consider reshuffling them during the breaks, as this can boost creativity.

## Rule 5: Consider Logistics, Facilities, and How to Record Outcomes

In advance, arrange to view the room that you will be using for the interactive workshop. The physical space(s) and equipment will influence the activities you run. Find out:

- Are there flip charts or blank walls for recording your participants' ideas? If not, do you need to order equipment such as foam boards?
- Can you rearrange the tables into small groups, or are there additional rooms available?

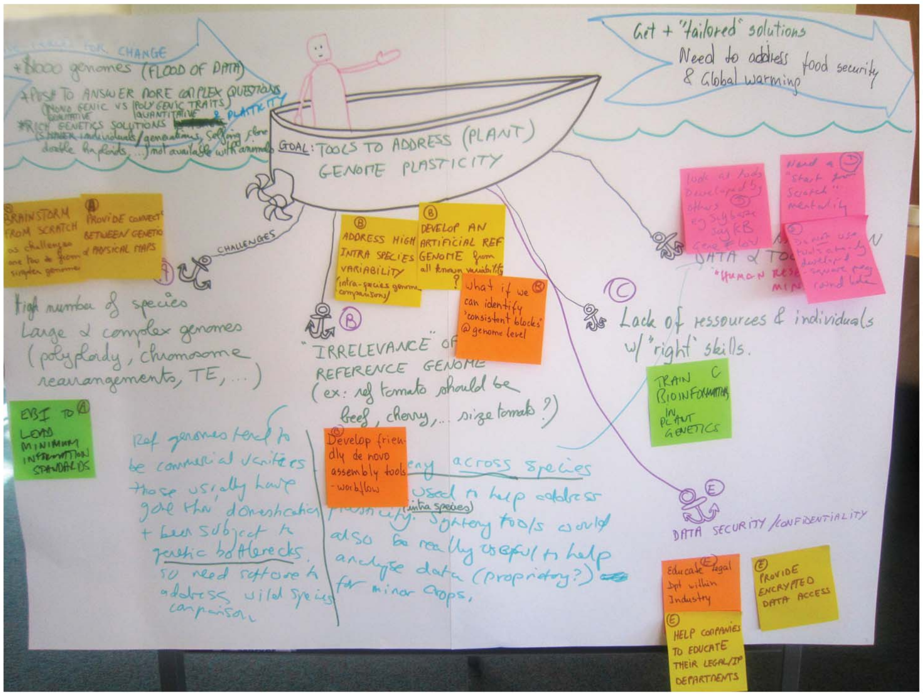



**Figure 2. Example of a workshop artefact: The output of the ''Speed Boat'' activity.** The aim of this activity is to identify improvements that need to be made, for instance, to a product or service. The boat and anchors are drawn on paper as a template before the workshop. During the activity, the groups add their ideas in pen: they write the goal of the workshop on the boat and the challenges to achieving this goal by the anchors. We also include ''positive forces for change''—things that are moving the project towards the goal—as ''wind arrows'' flanking the boat. The sticky notes have been added after the activity by the facilitator during the presenting-back stage and group discussion. The sticky notes have been labelled with the letters A to E for reference; note that a labelling scheme may be helpful for the analysis and report. This activity was adapted from p. 206 in [5]; also watch this video for more hints: http://www.youtube.com/watch?v=xwVbcioYvdM.
doi:10.1371/journal.pcbi.1003485.g002

- Is there audio-visual equipment, such as microphones for giving activity instructions?
- Is there a projector for electronic presentations of the activity instructions and templates?
- Are there areas for circulating during breaks?
- Can you arrange refreshments and/or catering?

The groups' outputs are called ''artefacts'' (see Figure 2 and http://www.youtube.com/watch?v=xwVbcioYvdM for an example). Artefacts could be whiteboards, flipchart paper sheets, drawings and/or sketches, canvasses (such as for the ''canvas sort'' activity in [2]), or 3D objects (such as boxes, as in the ''design the box'' activity in [5]). It is wise to take photographs of all artefacts on the day; we have found that paper is easily lost! Recording video summaries of the groups presenting back may also help clarify the information captured. Note that you will need to obtain consent from participants for video footage or photographs.

From our experience, interactive workshops are not suitable for remote participation, such as via webcasting, because the collaborative aspects of the group work cannot be easily shared remotely. If input from participants in distinct locations is necessary, consider running multiple smaller interactive workshops in different places.

## Rule 6: Plan the Agenda

You will need a minimum of 2–3 hours to run an interactive workshop but, ideally, a full day. For external participants, also consider the impact of travel arrangements on timing.

Ensure the agenda balances different types of activities, such as individual, paired, and collective tasks. Start with a hands-on activity as soon as possible and keep electronic-based presentations to a minimum—these can sap the creative atmosphere. Bear in mind the phases of a creative workshop: ''opening'' (generating ideas), ''exploring'' (experimenting with the ideas, finding patterns), and ''closing'' (evaluating, deciding, and listing actions) [5]. To conclude each activity, ensure you schedule enough time for presenting back to the group and discussing outcomes of each activity. This shows participants that their feedback and participation matters, and it provides an opportunity for clarification. Where possible, we use visual agendas because they set the participants' expectations of the day and prepare them for the interactive nature of the workshop (Figure 3).

Interactive workshops can be demanding. To maintain energy levels, plan breaks, provide refreshments, and aim to finish early. You may need to schedule the high priority activities earlier in the day in case enthusiasm wanes as you proceed. Alternatively, if you have participants who have travelled long distances and/or from different time zones, schedule the most important activity for later in the day.

Remember: planning is important, but you will need to be flexible with the timing on the day, for instance, if an activity is completed early or it overruns.

## Rule 7: Market Your Interactive Workshop As a Networking Opportunity

Networking is a great way to incentivise participation in a workshop with external candidates. Send named invitations and

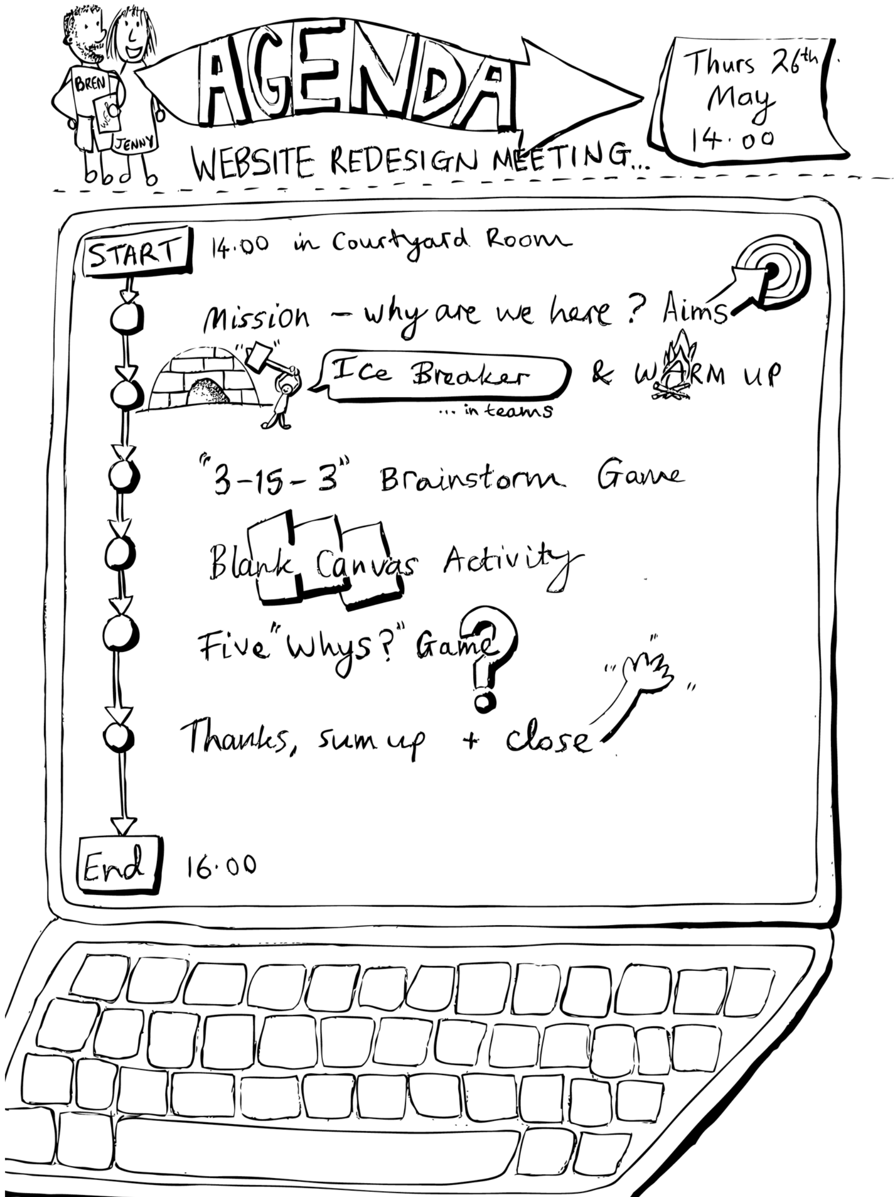



**Figure 3. Example of a visual agenda.** This was created for an internal workshop held to gather ideas for a website redesign process. Visual agendas are useful for setting the creative tone needed for successful interactive workshops.
doi:10.1371/journal.pcbi.1003485.g003

advertise the other organisations that will be present. You may wish to encourage the participants to network after the event by sharing their contact details. This can also be a useful contact list for your own future events. Remember to always get consent for how you intend to use contact information.

To encourage socialising at the interactive workshop, consider using games such as the "Low-tech Social Network" or "Show and Tell" [5]. We also recommend using name badges.

## Rule 8: Get the Best from Your Participants

As an interactive workshop leader, you need to encourage people to work together in a short space of time. We recommend displaying a table plan near the entrance to help participants settle quickly. We also try to give the groups amusing names to help set the creative tone and/or workshop theme.

Get the participants to introduce themselves, for example, by saying where they work, what they do, and why they chose to attend. This can be done for the whole room or group by group if it is a very large workshop. We find it is important for participants to know who is in the room before they will be comfortable sharing their ideas. After this, use "warm-up" and/or "ice-breaker" activities to stimulate collaborative working; for ideas, see [7].

For some activities, you may wish to have an example artefact pre-prepared, perhaps using a "toy" example. This can be more instructive than a list of instructions. Also have a template prepared for recording a synthesis of the presenting back (Figure 1).

Reflective personality types may find "on the spot" thinking uncomfortable [8], so provide information about the workshop aims in advance to help them prepare. On the day, using what we call a "creative silence" approach, where participants brainstorm ideas on sticky notes individually before sharing with the group, can help generate ideas and ensure that everyone participates.

If you have an international audience or diverse disciplines represented at your interactive workshop, you may need to do some background research. For example, some cultures have a tendency to be more reserved, whilst others may "warm up" quickly and may even need "cool down" activities to get the best results.

We find that details get noticed by the participants, so having creatively designed placeholders, good quality refreshments, a tidy and bright room, etc. make for a positive experience. The participants usually appreciate this and will in turn give more energy to taking part in the activities.

## Rule 9: Follow Up with Your Facilitators and Create a Post-Workshop Report

Immediately after the interactive workshop, ask your facilitators for their top three findings on the outcomes, and their main feedback on how the workshop went as a whole. It is best to do this face-to-face, so that you can reach a consensus quickly (because you will all be tired!). At the planning stage, schedule a meeting to do a more comprehensive analysis of the findings the day after the workshop. During the analysis you will need to synthesise and summarise the information on the artefacts.

You will probably need to present these outcomes in a report for whoever requested (and/or funded) the workshop. The executive summary is key because it highlights the main findings. We also suggest including the methods used for each game so that the reader can see how your results were gathered. Include quotes from participants to emphasise the results. Present your findings visually wherever possible—for instance, use graphs, charts, and photographs. Highlight any patterns that you identify and consensus from the delegates (e.g., "80% chose A, 20% chose B").

## Rule 10: Follow Up with Your Participants

Ask for feedback on the interactive workshop format: this is the best way to learn what works, and it shows your participants that you value their input. It is easiest if you can get the participants to fill out a short survey, ideally in the final coffee break so they complete it before they leave. Alternatively, you could send a thank you note to each participant with a link to a short electronic survey.

If the workshop was with stakeholders in a project (as opposed to users), then you may wish to share a report of the findings and a list of actions. It is important to do this soon after the event to maintain momentum. To strengthen the collaborations made with external participants, we have often arranged talks and outreach events at their institutes. Often the initial workshop gives us an insight into their requirements, which allows us to tailor activities to their needs.

Additionally, you may wish to contact participants for future projects (if you have obtained their consent for this). For instance, we ran a user-experience–based interactive workshop for a European Bioinformatics Institute (EBI) service, and once we had developed our ideas further, we invited the participants back to take part in usability testing of mock-ups of interface designs. We also asked if their colleagues would like to take part. This led to a day of testing with new participants and resulted in new collaborations. Networking with the participants is a good way to find opportunities for outreach, collaboration and future interactive workshops.


## Acknowledgments

Thanks to the EMBL-EBI Industry programme for granting us permission to use their Speed Boat activity chart. We also thank all of our previous interactive workshop participants and Francis Rowland (User Experience Designer) at EMBL-EBI, who helped edit this article.


### Further Reading

- Stephenson J, Galloway A (2012) The secrets to workshop success. Free e-book. ISBN: 978-87-403-0174-8. Available: http://bookboon.com/en/the-secrets-to-workshop-success-ebook. Accessed 14 August 2013.
- Sibbet D (2010) Visual Meetings: How graphics, sticky notes and idea mapping can transform group productivity. New Jersey: John Wiley & Sons.

## References


1. Koloski B (2012) Don't Have a Meeting, Throw a Workshop. UX Magazine, Article No. 852. Available: http://uxmag.com/articles/don%25E2%2580%2599t-have-a-meeting-throw-a-workshop. Accessed 14 August 2013.
2. de Matos P, Cham JA, Cao H, Alcántara R, Rowland F, et al. (2013) Enzyme Portal: a case study in applying user-centred design methods in bioinformatics. BMC Bioinformatics 14: 103. doi:10.1186/1471-2105-14-103
3. Fetzer J (2009) Quick, efficient, effective? Meetings! Anal Bioanal Chem 393(8): 1825–1827. doi:10.1007/s00216-009-2645-8.
4. Pruitt J, Adlin T (2006) The persona lifecycle: keeping people in mind throughout product design. New York: Morgan Kaufmann.
5. Gray D, Brown S, Macanufo J (2010) Game storming: A Playbook for Innovators, Rule-breakers and Changemakers. California: O'Reilly Media.
6. Rhizome. Guide to Facilitation Tools. Available: http://rhizomenetwork.files.wordpress.com/2010/12/facilitation_tools.pdf. Accessed 14 August 2013.
7. Seeds for Change. Facilitation Tools for Meetings and Workshops. Available: http://seedsforchange.org.uk/tools.pdf. Accessed 14 August 2013.
8. Briggs Myers I, Myers P (1995) Gifts Differing: Understanding Personality Type. Second edition. California: Davies-Black Publishing.

Editorial

# Ten Simple Rules for Organizing a Virtual Conference—Anywhere

**Nelson N. Gichora[1,2], Segun A. Fatumo[3], Mtakai V. Ngara[1,2], Noura Chelbat[4,5], Kavisha Ramdayal[6], Kenneth B. Opap[7], Geoffrey H. Siwo[8], Marion O. Adebiyi[3], Amina El Gonnouni[5], Denis Zofou[9], Amal A. M. Maurady[5], Ezekiel F. Adebiyi[3], Etienne P. de Villiers[10,11], Daniel K. Masiga[12], Jeffrey W. Bizzaro[13], Prashanth Suravajhala[13,14], Sheila C. Ommeh[10]*, Winston Hide[6,15]**

**1** International Institute of Tropical Agriculture, Nairobi, Kenya, **2** Faculty of Life Sciences, The University of Manchester, Manchester, United Kingdom, **3** Department of Computer and Information Sciences, Covenant University, Ota, Nigeria, **4** Institute of Bioinformatics, Johannes Kepler University, Linz, Austria, **5** Moroccan Society for Bioinformatics Institute, Morocco, **6** South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa, **7** University of Cape Town, Cape Town, South Africa, **8** University of Notre Dame, South Bend, Indiana, United States of America, **9** Biotechnology Unit, University of Buea, Buea, South West Region, Cameroon, **10** International Livestock Research Institute, Nairobi, Kenya, **11** Biosciences Eastern and Central Africa, Nairobi, Kenya, **12** International Center of Insect Physiology and Ecology, Nairobi, Kenya, **13** Bioinformatics Organization, Hudson, Massachusetts, United States of America, **14** Bioinformatics Team, Center for Development of Advanced Computing, Pune University Campus, Pune, India, **15** Harvard School of Public Health, Boston, Massachusetts, United States of America

The *First African Virtual Conference on Bioinformatics 2009* (AFBIX09) [1], organized by the Bioinformatics Organization [2] and the International Society for Computational Biology Student Council's Regional Student Groups of Africa and Morocco (ISCBSC RSG-Africa and RSG-Morocco) [3] received support from the African Society for Bioinformatics and Computational Biology (ASBCB) [4]. The aim was to provide students and scientists in the bioinformatics and computational biology fields a chance to network through a unique platform conceptualized as "hubs." These hubs then gave participants the opportunity to foster both physical and virtual interactions as well as develop collaborations, irrespective of geographical location.

Virtual conferencing may prove to be an effective low-cost strategy for conveying bioinformatics and computational biology education to African scientists who otherwise would be deprived of the opportunity. Unlike conventional conferences, virtual conferencing permits the involvement of a greater number of participants who would otherwise be unable to participate in events of this breadth owing to (1) limited travel fellowships, if any; (2) lack of time to travel to distant conference locations; and (3) insufficient accommodation and subsistence funds. These factors apply in general to the post-/undergraduate student community and especially to the target audiences that reside in developing countries. Minimizing the requirement to travel also means that the availability of invited speakers is greatly increased, improving the chances of attracting highly relevant and high-impact presenters.

Through the use of video conferencing software, virtual conferences are able to provide an accessible and cost-effective alternative to real time conferences while retaining the key benefits presented by an on-site conference, such as learning opportunities, sharing of ideas, and networking. The use of inexpensive "commodity off-the-shelf" (COTS) technologies permit anyone with an Internet connection, Web cam, and headset to give and/or attend a presentation. According to Andrew Sage, Cisco Systems' vice president for marketing, virtual conferences "can live on long after the physical booths have been torn down," while content continues to be viewed in a dedicated virtual environment by many people, even after the conclusion of the event [5].

At the Fall Joint Computer Conference on December 9, 1968, Douglas Engelbart presented, among other innovations, a virtual conferencing system that utilized the broadcast of computer monitor video as well as presenter audio and video [6]. This "expensive approach" has involved traditional video conferencing and technologies such as the Access Grid [7], which have been viable options for the most affluent regions of the world, but the approaches mentioned here are broad enough to be used in both developed and undeveloped environments.

The conference was set up as a series of virtual hubs defined as a group of ten or more persons in one location. Each hub consisted of a computer attached to a Web cam and speakers with a stable Internet connection. The hub activities and the interaction with other hubs were coordinated by persons within the locality.

Speakers within faculty and industry were identified on the basis of their expertise or involvement and relevance to the research topics covered by the virtual conference. There were a total of 16 speakers and out of these, four were keynotes divided between 2 days and four sessions. In addition, there were five invited speakers and three oral presentations selected from 12 submitted abstracts. The rest of the abstracts were presented as posters during break sessions. There were tutorials, relevant discussions from senior faculties, as well as welcome and closing statements from AFBIX09 organizers.

The conference was 19 hours long and was held over 2 days. The first day consisted

**Citation:** Gichora NN, Fatumo SA, Ngara MV, Chelbat N, Ramdayal K, et al. (2010) Ten Simple Rules for Organizing a Virtual Conference—Anywhere. PLoS Comput Biol 6(2): e1000650. doi:10.1371/journal.pcbi.1000650

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** February 26, 2010



**Funding:** SANBI hub was funded in part by the South African Medical Research Council (WAH). ILRI hub was funded in part by the International Livestock Research Institute and the Biosciences Eastern and Central Africa under a SIDA (Swedish) grant. Covenant University hub was funded by a special grant from the Covenant University Center for Research and Development (CUCERD). University of Notredame hub was supported by the Eck Institute for Global Health and the Notre Dame Bioinformatics Core. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: s.ommeh@cgiar.org

of 8 hours, tailored to accommodate time zone differences between each of the participating hubs. This was inclusive of 100 minutes of break time divided between two 20-minute coffee sessions concurrently spent on poster presentations, with an hour on a lunch break and 20-minute welcome speech. The second day consisted of an 11-hour program including one 20-minute coffee and poster session, 40-minute lunch break, and 30-minute vote of thanks and closing remarks.

The following ten simple rules are derived from experiences gained while organizing AFBIX09. We propose these as reference material to those intending to plan for similar events, with particular emphasis on resource-constrained communities.

## Rule 1: Address time zone differences: timing is everything.

Allow between 6 to 9 months before the conference to permit (1) administrators in the respective virtual hubs a sufficient amount of time to finalize their decisions regarding presentation and/or attendance time slots (relative to time zones) and (2) IT departments' confirmation for the provision of necessary support, amongst other logistics, for the designated event times. The organizing committee should agree on a conference schedule that will be suitable for the time zones of all participating groups.

It is effective to create a proposed conference program for all participating groups in their local time zones to avoid confusion. Once established, it is then crucial to conduct tests of the proposed times precisely as scheduled, weeks before the actual event, to ensure the reliability of the conference program and to identify problems that could arise.

## Rule 2: Test the available resources: to ensure that you are able to host the conference.

Ensure the availability of (1) a stable Internet connection; (2) a computer installed with the required video-conferencing software; (3) reliable audio speakers that have been tested for audio clarity; (4) adequate screen resolution for the capabilities of the network; and (5) a public-address system (i.e., video camera and projector connections). There should be adequate lighting for the conference hall to avoid glare or other aspects of poor visibility. Another useful resource is a standby computer assigned to the hub-coordinator with a communication application/device, such as a VoIP service, in place to ensure synchronous coordination of the proceedings with other participating hubs.

As an illustration, the last point was particularly useful in an instance where two of the participating hubs during the conference experienced network downtime, cutting off real-time presentations. Before the restoration of network connection, the respective hub coordinators had to inform the other hubs of their downtime and continually synchronize conference activities.

## Rule 3: Manage bandwidth usage: to safeguard against conference interruptions.

It is critical and advisable to make sure your organizations' IT personnel are able to allocate sufficient bandwidth to the virtual conference, to avoid disruptions of live presentations (especially in organizations where network resources are shared). Alternatively, if a group of 10 or more participants are registered for the conference, it is advisable that these individuals form an independent virtual hub to save on bandwidth usage. This approach will reduce the number of Internet connections being used and thus the potential complications for your virtual conference while allowing other users an equally reliable functioning network.

## Rule 4: The concept of virtual hubs: makes registration and participation simpler.

Distribute the virtual conference registration fee across all participating hubs and participants [8–12]. Cumulative hub payments ensure a reduced registration fee for the individual participant. Hubs provide local expertise and relevant local advertising for the conference. These "front porch" gathering sites compensate for some of the personal interaction that can be missing from virtual conferences. The use of virtual hubs as "conference nodes" tends to increase impact by providing access for those without the equipment and also traditional face-to face interaction. Hub participants can also share traditional meeting activities such as enjoying a meal together.

## Rule 5: Prerecord presentations: to gear-up if streaming video fails for any reason.

There is a wide range of software available to get connected virtually (e.g., WebEx, Netviewer, Adobe Connect, etc.), however all available Internet systems are subject to bandwidth limitations and resulting congestion. It is therefore advisable that presentations be prerecorded and in no less than 2 weeks before the conference, in order to permit time for the recordings to be edited or redone, if necessary. Prerecorded presentations can then be hosted via the conference Web sites, making them available to the participating groups in an agreeable video format and in good time to conduct/resolve software compatibility concerns. Moreover, this allows the participants a chance to become familiar with the conference content and to play back presentations containing key concepts/information. The use of prerecorded presentations compensates for slow and unreliable networks and even intermittent electrical outages (e.g., when two of the aforementioned hubs experienced connectivity problems, they resorted to projecting prerecorded presentations to the participants in their respective hubs, and when this was resolved they were able to join the live Q&A sessions). Alternatively, if the network problems are not restored in time, the narrator can then appear online after the prerecorded presentation to answer questions in real time or to take questions via a text-based chat system.

## Rule 6: Allocate time for presenter orientation: to ensure glitch-free schedule compliance.

Keynote and invited presenters should become familiar with the designated software, preferably a month before the conference. This will enable them to get acquainted with the software while allowing them to prerecord their own presentation at their convenience. Recorded presentations should then be sent to the conference host, who should test and archive all recordings before use if/when the scheduled presenter is absent at the time of his/her presentation.

## Rule 7: Establish dedicated virtual interaction rooms (e-lobbies): to ensure a practical platform for participant Q&A and networking.

Each participating hub should have at least one person responsible for the collection and consolidation of all participant questions or answers from that hub. This consolidation avoids redundancy while saving time and kilobytes. Alternatively, the designated person could verbally relay the questions to the presenters on behalf of the hub to ensure clarity. This approach is especially applicable in cases where one of the hubs is in a country where the language of instruction is not the one adopted for the conference. The availability of "e-lobbies" will permit the comfortable virtual interaction of participants with similar research interests during virtual poster sessions and/or coffee breaks.

## Rule 8: Troubleshoot technical glitches: to equip yourself for any foreseeable challenges.

Identify at least one person per hub to coordinate the technical set-up of the conference venue and to ensure, well in advance, that all technical equipment and relevant software are available and functioning properly.

## Rule 9: Get motivated… It's the key to your success.

It is crucial to be able to set and meet your deadlines/milestones through adequate time management, hub organization, etc. Besides this, involve people who are inspired, willing, and passionate to organize the conference. Encourage participants in different hubs to take photos throughout the event. The effects of team building last long after the conference, and encouraging participation results in leadership development. Plus, the managerial skills developed play an enormous part in the success of the conference.

## Rule 10: Participant feedback: useful for future reference.

At the conclusion of the conference, be sure to request feedback from the participants to be able to identify any faults or errors that can then be addressed in future events. Make sure to have all questions that were raised during the presentations and their corresponding answers available online to all participants including photos taken during the event. Aside from having this information on record, it will help sustain communication even after the virtual conference has been concluded.

The recorded videos and presentations have been made available through Bioinformatics.Org and hyperlinked on the wiki page at http://www.bioinformatics.org/wiki/Afbix09. Bioinformatics.Org seeks the opinions of the community via online polls. Blogging was not implemented in this conference, but we envisage that the online educational system operated at Bioinformatics.Org could be utilized for that in the future.

## Valuable Lessons

Overall, what worked included prerecording the presentations, which were of great assistance when streaming video failed. Use of a chat facility (e.g., Skype) was key in coordinating hub activities during the course of the conference as some of the participating hubs experienced connectivity problems and had to synchronize their prerecorded presentation with the live presentations being viewed by other hubs.

What didn't work included disruption in the streaming video, which was a major drawback, and resulted in most hub coordinators relying on prerecorded videos of the conference presentations. Virtual interaction rooms (e-lobbies) were not effectively utilized as earlier anticipated; this was in contrast to the hub level where participants were able to effectively interact. It would be useful to set up subcommittees in order to deal with conference requirements as they arise. These include technical committees, fundraising committees, and scientific committees among others. It is also important for all committee members to meet regularly with the frequency of meetings increasing as the conference start date draws near.

## Impact on Science in Africa

The novel idea of virtual hubs through e-conferencing was pioneered in AFBIX09. With a stable Internet connection, the maximum number of participants at any conference is dependent on whether future conferences will adopt the concept of virtual hubs. This means that the audio-visual facilities in each hub and sitting space should dictate the maximum number of persons in one hub as compared to the single user participation option. Depending on the choice of the video-conferencing software and the maximum number of connections it can allow at a given time, this value can be translated to hubs. Therefore the number of participants that can attend a virtual meeting will depend on the number of formed hubs and consequently, the maximum capacity of each hub, which may translate to thousands of participants. A new high bandwidth optical fiber cable is being laid around the coast of Africa with bandwidth improvements of 10–100 times expected around most places in Africa. This development should greatly affect future virtual activities within the continent. The African Virtual Conference on Bioinformatics (AFBIX), which was a hybrid between a normal and virtual conference, has had a large impact in the field and consequently there are plans to hold it biennially. This has impacted greatly on ISCB Regional students groups (see below) as well as other spin-off conferences such as the Indian Virtual Conference on Bioinformatics (Inbix10, http://www.bioinformatics.org/wiki/Inbix10).

In terms of participants, the Regional Student Group (RSG)-Moroccan hub had a total of 12 attendees for the AFBIX09, which enabled RSG-Morocco to develop a working relationship/collaboration with the Institut Pasteur de Tunis in Tunisia. The presentations made during the conference sparked discussions between students and scientists touching on the various topics covered, leading to the forging of new ideas on possible bioinformatics projects to undertake.

The RSG-Africa-Southern Africa hub attracted on average ten attendees for the 2 days. The hub was faced with technical issues that affected the quality of the presentations. Although overall, the attendees benefited greatly and called for improvement of future conferences.

The RSG-Africa-Eastern Africa hub attracted a total of 25 attendees as a result of a collaborative effort between the Biosciences East and Central Africa (BecA), who funded all of the students, and the International Livestock Research Institute (ILRI), who provided conferencing facilities gratis. The success of AFBIX09 prompted members to come up with plans to start collaborative bioinformatics projects between RSG-Africa-Eastern Africa and other RSGs, organizations, or institutes that will enable greater collaborations in research and training. The hub also established contacts with RSG-India, which has experience in virtual collaborative bioinformatics projects.

The RSG-Africa-Western Africa hub had a total of 17 attendees. The conference provided a platform for forging collaboration between the biological sciences and computer science departments at Covenant University, which acted as the hub for the conference. The conference attracted key administrators in their institute, including the vice chancellor, and this was a great boost for the students' group of West Africa.

The University of Notre Dame had an average range of eight to 20 attendees.In addition, three other faculties participated in the conference. This was a sure venue to foster collaboration with other students in developing countries.

The total number of participants, including speakers, organizers, and single user participants was close to 100. In conclusion, although several challenges were experienced, AFBIX09 has established a foundation for future virtual conferences.


## Acknowledgments

We acknowledge the contributions made by all participating institutions and organizations: The Regional Student Group of Africa (RSG-Africa) and Regional Student Group of Morocco (RSG-Morocco), both of which are affiliates of the International Society for Computational Biology Student Council (ISCBSC). The five independent virtual hubs: International Livestock Research Institute/Biosciences East and Central Africa (ILRI/BecA), Kenya [5]; South African National Bioinformatics Institute (SANBI) MRC/UWC/SANBI Bioinformatics Capacity Development Unit, University of the Western Cape, South Africa [6]; Moroccan Society for Bioinformatics Institute (SMBI), Morocco [7]; Covenant University, Nigeria [8]; University of Notre Dame, USA [9] and the Bioinformatics Organization, USA [1] as the host. We thank Sonal Patel (ILRI) and Dale Gibbs and Mario Jonas (SANBI) for their kind voluntary assistance.


## References


1. African Virtual Conference on Bioinformatics 2009 (AFBIX09). Available: http://wiki.bioinformatics.org/Afbix09.
2. Bioinformatics Organization. Available: http://www.bioinformatics.org/.
3. International Society for Computational Biology, Student Counci. Available: http://www.iscbsc.org/.
4. African Society for Bioinformatics and Computational Biology (ASBCB). Available: http://www.asbcb.org/.
5. Link to "Virtual conferences home advantage" in BusinessWeek. Available: http://www.businessweek.com/technology/content/may2008/tc2008054_560356.htm.
6. Engelbart D, English W (1968) A research center for augmenting human intellect. AFIPS Conference Proceedings of the 1968 Fall Joint Computer Conference; December 1968; San Francisco. Volume 33. pp 395–410.
7. Reed D, Mendes C, Lu C, Foster I, Kesselman K (2003) The Grid 2: blueprint for a new computing infrastructure-Application Tuning and adaptation. 2nd edition. San Francisco: Morgan Kaufman. pp 513–532.
8. International Livestock Research Institute (ILRI). Available: http://www.ilri.org/.
9. South African National Bioinformatics Institute (SANBI). Available: http://www.sanbi.ac.za/.
10. Covenant University. Available: http://www.covenantuniversity.com/.
11. University of Notre Dame. Available: http://www.nd.edu/.
12. Moroccan Society for Bioinformatics Institute (SMBI). Available: http://www.smbi.ma/.

Editorial

# Ten Simple Rules for Chairing a Scientific Session

**Alex Bateman[1], Philip E. Bourne[2]***

1 Wellcome Trust Sanger Institute, Cambridge, United Kingdom, 2 Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, California, United States of America

Chairing a session at a scientific conference is a thankless task. If you get it right, no one is likely to notice. But there are many ways to get it wrong and a little preparation goes a long way to making the session a success. Here are a few pointers that we have picked up over the years.

## Rule 1: Don't Let Things Overrun

Probably the main role of the session chair is to keep the meeting running on time. Time is a strange and elastic concept when people are under pressure. Some speakers will talk much faster than normal and finish a talk in half the expected time. Others will ramble on without knowing that time is running out and they have only just finished their introduction. Timing is important to ensure that a meeting runs smoothly. Delegates should leave the session at just the right time so that lunches are still fresh, bars still open, etc. Timing is particularly acute if there are multiple parallel sessions and delegates would want to switch between talks in different sessions.

## Rule 2: Let Your Speakers Know the Rules

A session will run more smoothly if you let all the speakers know how you plan to run your session. This could be done by e-mail before the event or you might want to gather up the speakers just before the session. Reminding them how much time they have to speak, how much time to allow for questions, and how you will let them know time is up will stop confusion later on. Beyond the rules, encourage speakers to review what others in the session will say. The less redundancy, the better the session will be for everyone, including the chair.

## Rule 3: Be Prepared to Give a Short Introduction

Be prepared to give a short introduction to the session, and, of course, introduce yourself as well. Be sure to review the abstracts of the talks and then give a succinct summary of what will be presented. It is your job to excite people at the session and have them stay in the auditorium. Regarding the speakers, introduce each one before they begin, providing their background and highlighting their major accomplishments. Speakers love to be properly introduced and the audience likes to feel they know the person speaking. But for the sake of both the timing of the session and your speakers, do keep it brief. Are you expected to give any housekeeping messages or to remind people to switch off their phones? Allow time for that if so.

## Rule 4: Write Down the Actual Start Times of the Speakers

If you don't know what time a speaker started, it is difficult to know when to ask them to stop. So always write down the start and finish times of speakers throughout the session.

## Rule 5: Do Have a Watch

It sounds obvious, but it is very difficult to chair a session if you don't have a watch and don't know the time. Yes, one of us has done this! It is embarrassing to have to ask your neighbor for a watch. Actually, it is probably best to have two watches, just in case.

## Rule 6: Communicate How Much Time is Left to the Speaker

Letting the speaker know their time is up is crucial in keeping time. A simple sign held up at the right time is usually fine. Have one saying, "5 minutes to go" and another saying "time is up". Beyond that time, standing up on the stage is a good sign that the speaker should wrap up.

## Rule 7: Don't Be Afraid to Move on Without Questions

A good scientific session is characterized by a lively question and answer session. In fact, some speakers believe it is their right to expect to answer questions even after their allotted time is up. If you are running over time, you should not be afraid to move on to the next talk without questions. You will be more confident in enforcing this principle if you have warned the speaker beforehand that running over will require foregoing taking questions at that time. You can stay on schedule by diplomatically saying that the speaker will be happy to take questions at the break.

## Rule 8: Get to the Venue Early and Be Audiovisually Aware

Make sure to know where everything is, like pointers, microphones, projectors, and computers and who to turn to if it all goes wrong. It is worth checking that all these things work so that you can swiftly fix them yourself. Knowing ahead of time any unusual requests from speakers to show movies and sound clips requiring special attention. Be sure the venue supports the needs of speakers. If not, let them know before they get to the venue. If each speaker is expected to load their presentation on a single computer associated with the podium, allow time for that and have the speaker run through their slides to be sure everything is working properly.

## Rule 9: Prepare Some Questions in Advance

It can take an audience a few seconds to digest the contents of a talk and think of questions. So, it is always good to have one or two ready to ask. These can be

**Citation:** Bateman A, Bourne PE (2009) Ten Simple Rules for Chairing a Scientific Session. PLoS Comput Biol 5(9): e1000517. doi:10.1371/journal.pcbi.1000517

**Published** September 25, 2009



**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: bourne@sdsc.edu

prepared beforehand from the abstracts and supplemented from ones that occur to you during the talk. This is a very good reason for paying attention during the talk. Also, it is worth thinking of one or two general purpose questions such as "What do you plan to do next?"

## Rule 10: Keep Control of the Question and Answer Sessions

It is difficult for the session chair to keep things on time if the speaker is in control of taking questions. Make sure you are the one who selects the next questioner. Also, be prepared to step in if the speaker and questioner are getting into a long-winded, technical discussion.

Hopefully with a bit of preparation and a little luck, you will get through the ordeal of chairing a scientific session unscathed. And remember, if no one thanks you, you have probably done an excellent job.

Editorial

# Ten Simple Rules for Organizing a Scientific Meeting

**Manuel Corpas[1], Nils Gehlenborg[1,2], Sarath Chandra Janga[3], Philip E. Bourne[4]***

**1** European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, **2** Graduate School of Life Sciences, University of Cambridge, Cambridge, United Kingdom, **3** Medical Research Council–Laboratory of Molecular Biology, University of Cambridge, Cambridge, United Kingdom, **4** Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, California, United States of America

Scientific meetings come in various flavors—from one-day focused workshops of 1–20 people to large-scale multiple-day meetings of 1,000 or more delegates, including keynotes, sessions, posters, social events, and so on. These ten rules are intended to provide insights into organizing meetings across the scale.

Scientific meetings are at the heart of a scientist's professional life since they provide an invaluable opportunity for learning, networking, and exploring new ideas. In addition, meetings should be enjoyable experiences that add exciting breaks to the usual routine in the laboratory. Being involved in organizing these meetings later in your career is a community responsibility. Being involved in the organization early in your career is a valuable learning experience [1]. First, it provides visibility and gets your name and face known in the community. Second, it is useful for developing essential skills in organization, management, team work, and financial responsibility, all of which are useful in your later career. Notwithstanding, it takes a lot of time, and agreeing to help organize a meeting should be considered in the context of your need to get your research done and so is also a lesson in time management. What follows are the experiences of graduate students in organizing scientific meetings with some editorial oversight from someone more senior (PEB) who has organized a number of major meetings over the years.

The International Society for Computational Biology (ISCB) Student Council [2] is an organization within the ISCB that caters to computational biologists early in their career. The ISCB Student Council provides activities and events to its members that facilitate their scientific development. From our experience in organizing the Student Council Symposium [3,4], a meeting that so far has been held within the context of the ISMB [5,6] and ECCB conferences, we have gained knowledge that is typically not part of an academic curriculum and which is embodied in the following ten rules.

## Rule 1: The Science Is the Most Important Thing

Good science, above all else, defines a good meeting; logistics are important, but secondary. Get the right people there, namely the best in the field and those who will be the best, and the rest will take care of itself. When choosing a topic for your conference, map it to the needs of your target audience. Make sure that you have a sufficiently wide range of areas, without being too general. The greater the number of topics covered, the more likely people are to come, but the less time you will have to focus on particular subject matter. Emerging areas can attract greater interest; try to include them in your program as much as possible; let your audience decide the program through the papers they submit to the general call for papers. This can be done with broad and compelling topic areas such as "Emerging Trends in …" or "New Developments in …".

## Rule 2: Allow for Plenty of Planning Time

Planning time should range from nine months to more than a year ahead of the conference, depending on the size of your event. Allow plenty of time to select your meeting venue; to call for, review, and accept scientific submissions; to arrange for affordable/discounted hotel rooms; to book flights and other transportation options to the conference. Having outstanding keynote speakers at your event will also require you contact them months in advance—the bigger the name, the more time is required.

## Rule 3: Study All Potential Financial Issues Affecting Your Event

Sponsors are usually your primary source of funds, next to the delegates' registration fees. To increase the chances of being sponsored by industry, write them a clear proposal stating how the money will be spent and what benefits they can expect to get in return. You may also want to reserve a few time slots for industry talks or demos as a way of attracting more sponsors, but be wary that the scientific flavor of the meeting is not impacted by blatant commercialism. Make sure you first approach the sponsors that match your interest topics the closest. If they say they are not interested this year, keep their contact information, as they might be able to sponsor you in future events. Approach them early rather than later in any case. The cost of your conference will be proportional to the capacity of the venue; therefore, a good estimation of the number of attendees will provide you with a good estimate of your costs. You will need to include meals and coffee breaks together with the actual cost of renting your venue. Be aware that audiovisual costs can be additional as well as venue staff—look out for hidden costs. Aside from venue-related costs, additional expenditures might include travel fellowships, publication costs for proceedings in a journal, and awards for outstanding contributors. All these issues will determine how much you need to charge your participants to attend. Map all this out on a spreadsheet and do the math. Allow for contingencies, such as currency fluctuations and world-changing

**Citation:** Corpas M, Gehlenborg N, Janga SC, Bourne PE (2008) Ten Simple Rules for Organizing a Scientific Meeting. PLoS Comput Biol 4(6): e1000080. doi:10.1371/journal.pcbi.1000080

**Published** June 27, 2008



**Funding:** The authors have received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: bourne@scsd.edu

events that will impact attendance. For large meetings, consider insurance against such events. Starting with a template that others have used for previous similar conferences can be a big help.

## Rule 4: Choose the Right Date and Location

Your conference needs to be as far away as possible from established conferences and other related meetings. Alternatively, you may want to organize your event around a main conference, in the form of a satellite meeting or Special Interest Group (SIG). Teaming up with established conferences may increase the chances of attracting more people (especially if this is your first time) and also save you a great deal of administrative work. If you decide to do it on your own, you should consider how easy it is to travel to your chosen location, whether it has a strong local community in your field, and whether it has cultural or other tourist attractions. Inexpensive accommodation and airfares to your conference are always a plus.

## Rule 5: Create a Balanced Agenda

A conference is a place for people wanting to share and exchange ideas. Having many well-known speakers will raise the demand for your event (and the cost) but that has to be balanced with enough time for presentation of submitted materials. A mix of senior scientists and junior scientists always works for the better. Young researchers may be more enthusiastic and inspiring for students, while top senior scientists will be able to present a more complete perspective of the field. Allow plenty of time for socializing, too; breaks, meals, and poster sessions are ideal occasions to meet potential collaborators and to foster networking among peers.

## Rule 6: Carefully Select Your Key Helpers: the Organizing Committees

A single person will not have all the skills necessary to organize a large meeting, but the organizing committee collectively needs to have the required expertise. You might want to separate the areas of responsibilities between your aides depending on their interests and availability. Some potential responsibilities you might delegate are: 1) content and design of the Web site promoting the meeting; 2) promotion materials and marketing; 3) finance and fundraising; 4) paper submissions and review; 5) posters; 6) keynotes; 7) local organization; 8) program and speakers; 9) awards. Your organizing committee should be large enough to handle all the above but not too large, avoiding freeloaders and communication issues. It is invaluable to have a local organizing committee since they know local institutions, speakers, companies, and tourist attractions. Local organizations may also help you with administrative tasks; for example, dealing with registration of attendees and finding suitable accommodations around the venue.

## Rule 7: Have the Members of the Organizing Committees Communicate Regularly

It is good to have planning sessions by teleconference ahead of the meeting. As far as possible, everyone should be familiar with all aspects of the meeting organization. This collective wisdom will make it less likely that important issues are forgotten. The local organizers should convince everyone that the venue will work. Use these sessions to assign responsibilities ahead of the meeting. Tasks such as manning the registration tables, carrying microphones for attendees to ask questions, introducing sessions and speakers, checking presentations ahead of time, and having poster boards, materials to attach posters, etc., are easily overlooked. In short, good communication will lead to you covering all the little things so easily forgotten.

Good communication continues throughout the meeting. All organizers should be able to contact each other throughout the meeting via mobile phone and e-mail. Distribute to all organizers the names and contact information of caterers, building managers, administrative personnel, technicians, and the main conference organizer if you are having your event as part of another conference. Onsite changes that incur additional costs, however, should require the approval of a single, key organizer rather than all organizers operating independently of one another. This will ensure there are no financial surprises in the end. It is also important that you have a designated meeting point where someone from the organizing committee is going to be available at all times to help with problems.

## Rule 8: Prepare for Emergencies

Attendees need to be aware of all emergency procedures in terms of evacuation, etc. This should be discussed with the venue managers. All attendees should be reachable as far as possible during the conference. If an attendee has an emergency at home, his or her family should be able to reach them through the conference desk—mobile phones are not perfect after all.

## Rule 9: Wrap Up the Conference Properly

At the end of the conference, you should give credit to everyone who helped to make the event a success. If you have awards to present, this is the right time for the awards ceremony. Dedicate some time to thank your speakers and sponsors as well as everyone involved in the organization of the conference. Also collect feedback about the event from the delegates through questionnaires. This evaluation will help you to understand the strengths and weaknesses of your conference and give you the opportunity to improve possible future events. Have a party or some other event for all those organizing the conference.

## Rule 10: Make the Impact of Your Conference Last

Published proceedings are the best way to make the results of your conference last. Negotiate with journals far in advance of the conference to publish the proceedings. Make those proceedings as widely accessible as possible. Upload photos and videos of the event to the conference Web site and post the names of presenters who have received awards or travel fellowships. It is also a good idea to link the results of your evaluation to the Web site. Send one last e-mail to all delegates, including a summary of the activities since the conference and thanking them for their participation. This is particularly important if you are considering holding the conference again in future years, in which case include some information on your plans for the next event.

As always, we welcome your comments and experiences that you think would enrich these ten rules so that they might be useful to others. The comment feature now supported by this journal makes it easy to do this.

## Acknowledgments

We would like to acknowledge the International Society for Computational Biology (ISCB) for their support in the organization of the Student Council Symposiums, in particular BJ Morrison-McKay and Steven Leard. Thanks to Michal Linial and Rita Casadio (our liaisons at the ISCB Board of Directors), Burkhard Rost (the ISCB President), and all the ISCB Board of Directors for being so supportive of the work of the Student Council. We are also grateful to all the Student Council leadership and current and past Student Council members for their enthusiasm and hard (unpaid) work. You all have made the Student Council a great organization.

## References

1. Tomazou EM, Powell GT (2007) Look who's talking, too: Graduates developing skills through communication. Nat Rev Genet 8: 724–726. doi:10.1038/nrg2177.
2. The International Society for Computational Biology Student Council. Available: http://www.iscbsc.org. Accessed 22 April 2008.
3. Corpas M (2005) Scientists and societies. Nature 436: 1204. doi:10.1038/nj7054–1204b.
4. Gehlenborg N, Corpas M, Janga SC (2007) Highlights from the Third International Society for Computational Biology (ISCB) Student Council Symposium at the Fifteenth Annual International Conference on Intelligent Systems for Molecular Biology (ISMB). BMC Bioinformatics 8 (Supplement 8):I1.
5. Lengauer T, McKay BJM, Rost B (2007) ISMB/ECCB 2007: The premier conference on computational biology. PLoS Comput Biol 3: e96. doi:10.1371/journal.pcbi.0030096.
6. Third ISCB Student Council Symposium. Available: http://www.iscbsc.org/scs3 Accessed 22 April 2008.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks

**Adam Rule[1], Amanda Birmingham[2], Cristal Zuniga[3], Ilkay Altintas[4], Shih-Cheng Huang[4¤], Rob Knight[3,5], Niema Moshiri[6], Mai H. Nguyen[4], Sara Brin Rosenthal[2], Fernando Pérez[7], Peter W. Rose[4]***

**1** Design Lab, UC San Diego, La Jolla, California, United States of America, **2** Center for Computational Biology and Bioinformatics, UC San Diego, La Jolla, California, United States of America, **3** Department of Pediatrics, UC San Diego, La Jolla, California, United States of America, **4** Data Science Hub, San Diego Supercomputer Center, UC San Diego, La Jolla, California, United States of America, **5** Departments of Bioengineering, and Computer Science and Engineering, and Center for Microbiome Innovation, UC San Diego, La Jolla, California, United States of America, **6** Bioinformatics and Systems Biology Graduate Program, UC San Diego, La Jolla, California, United States of America, **7** Department of Statistics and Berkeley Institute for Data Science, UC Berkeley, and Lawrence Berkeley National Laboratory, Berkeley, California, United States of America

¤ Current address: Biomedical Informatics Graduate Program, Stanford University, Stanford, CA, United States of America
* pwrose@ucsd.edu





**Citation:** Rule A, Birmingham A, Zuniga C, Altintas I, Huang S-C, Knight R, et al. (2019) Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks. PLoS Comput Biol 15(7): e1007007. https://doi.org/10.1371/journal.pcbi.1007007

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** July 25, 2019



**Funding:** The authors received no specific funding for writing this article. Funding for the "Reproducible Research and Interactive Education - Application of Jupyter Notebooks" workshop was provided by the San Diego Supercomputer Center at UC San Diego.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

As studies grow in scale and complexity, it has become increasingly difficult to provide clear descriptions and open access to the methods and data needed to understand and reproduce computational research. Numerous papers [1–3], including several in the Ten Simple Rules collection [4,5], have highlighted the need for robust and reproducible analyses in computational research, described the difficulty of achieving these standards, and enumerated best practices. We aim to augment this existing wellspring of advice by addressing the unique challenges and opportunities that arise when using computational notebooks, especially Jupyter Notebooks, for research [6].

Reproducibility, the scientific standard that others should be able to recreate your results, requires at a minimum that "data and the computer code used to analyze [that] data be made available to others" [2]. Achieving even this minimum standard typically requires both machine-readable descriptions of the data, software, dependencies, and computational environment involved (for example, hardware or cloud configuration), as well as human-readable documentation describing how all these pieces fit together. Whereas analysts previously kept code, documentation, and results in separate files, they increasingly use computational notebooks such as Jupyter Notebooks and R Notebooks to both perform analyses and combine code, results, and descriptive text in a single "computational narrative" to be read and rerun by others [7,8]. This ability to combine executable code and descriptive text in a single document has close ties to Knuth's notion of "literate programming" [9] and has convinced many researchers to switch to computational notebooks from other programming environments. Jupyter Notebooks in particular have seen widespread adoption: as of December 2018, there were more than 3 million Jupyter Notebooks shared publicly on GitHub (https://www.github.com) [10], many of which document academic research [11].

The interactive and narrative nature of computational notebooks presents unique opportunities for performing and sharing computational research. With some forethought, they can provide not only richly detailed descriptions of analyses but also interactive computing

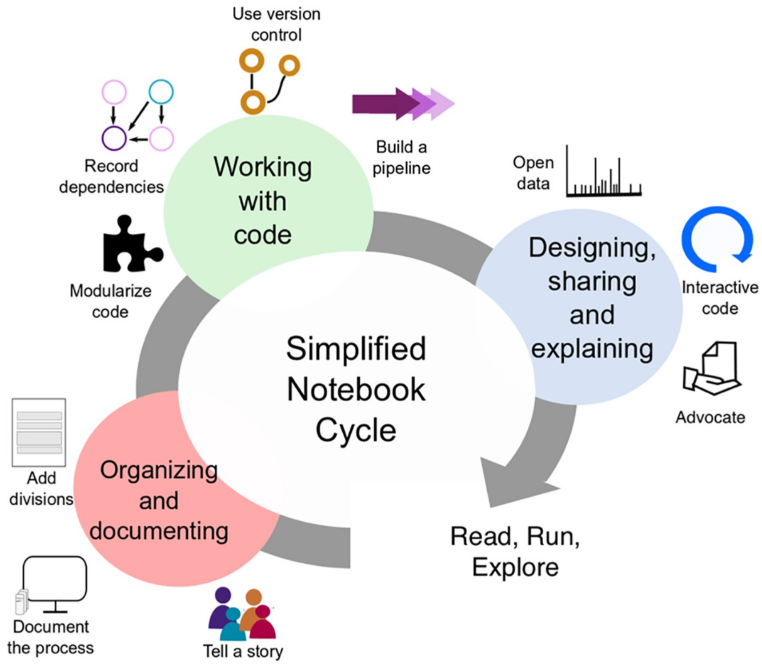



**Fig 1. Iterative workflow for applying the 10 simple rules to the creation of Jupyter Notebooks.** The cycle describes three overlapping phases of developing a well-documented and functional Jupyter Notebook. First, we organize and document the notebook (Rules 1–3). Second, the code is developed following the rules proposed here about quality standards (Rules 4–7). Finally, the notebook is made available, along with its data (Rule 8), in a manner encouraging public exploration and contribution (Rules 9–10).

https://doi.org/10.1371/journal.pcbi.1007007.g001

environments for replicating, exploring, and extending them. Yet, as with other computing environments, using notebooks for research requires special care. Interactively running and editing code in notebooks can delete key steps or introduce "hidden state" that confounds analyses and confuses readers [12]. Analyses documented in notebooks cannot be easily rerun if users do not first freeze their dependencies, share their data, and adequately describe their computing environment [13]. And many notebooks lack sufficient descriptive text to guide readers in using them [11,14].

The explosive growth of computational notebooks provides a unique opportunity to support computational research, but care must be taken when performing and sharing analyses in notebooks. Given these opportunities and challenges, we have compiled a set of rules, tips, tools, and example notebooks to help guide Jupyter Notebook authors. While we focus on a few core uses of Jupyter Notebooks observed in our own research, many of these rules can be applied to other computational notebooks and use cases. In Fig 1, we give a preview of the rules applied at different phases of the notebook development cycle. Whether you use notebooks to track preliminary analyses, to present polished results to collaborators, as finely tuned pipelines for recurring analyses, or for all of the above, following this advice will help you write and share analyses that are easier to read, run, and explore.

## Rule 1: Tell a story for an audience

One key benefit of using Jupyter Notebooks is being able to interleave explanatory text with code and results to create a computational narrative [7]. Rather than only keep sporadic notes, use explanatory text to tell a compelling story that has a beginning that introduces the topic, a middle that describes your steps, and an end that interprets the results. Describe not just what you did but why you did it, how the steps are connected, and what it all means. It is okay for your story to change over time, especially as your analysis evolves, but be sure to start documenting your thoughts and process as early as possible.

How you tell the story will depend on your goal and audience. Do you plan to share your notebook with a nontechnical colleague in your lab, analysts at another lab, readers of a particular journal, or the general public? You may need different kinds and levels of explanation for each audience. In any case, remember that your primary audience will most likely be your future self. Is your explanation clear enough that you will be able to understand and reproduce the analysis a month from now? People often overestimate what they will be able to remember in the future, so err on the side of overexplaining. If you won't be able to recreate your own analysis in the near future, how could anyone else?

## Rule 2: Document the process, not just the results

Computational notebooks' interactivity makes it quick and easy to try out and compare different approaches or parameters—so quick and easy that we often fail to document those interactive investigations at the time we perform them. Thus, the advice long provided regarding paper lab scientific notebooks becomes even more critical: make sure to document all your explorations, even (or perhaps especially) those that led to dead ends. These comments will help you remember what you did and why. You can always remove these comments later if turning your notebook into a pipeline (see Rule 7) or preparing to share it with a wider audience (Rule 1), who may prefer to see a concise presentation of results rather than a detailed lab notebook.

Many notebook users wait to add such explanatory text until the end of an analysis, after they have a solid result. Don't wait—by that point you may have forgotten why you chose a particular parameter value, where you copied a block of code from, or what you found interesting about an intermediate result. If you do not have time to fully document what you are doing or thinking in the moment, leave short descriptive notes to remind yourself what to add when you get to a good stopping point. While the code needed to reproduce the analysis may be automatically captured in your notebook, the reasoning and intuition may not. It is okay if the story in your notebook changes over time; you should still tell a story from the very beginning, even if you don't know the ending yet.

Clean, organize, and annotate your notebook after each experiment or meaningful chunk of work and do all your cleaning in the notebook. For example, when preparing to publish, avoid manually tweaking figures with desktop publishing tools and instead use plotting libraries with the notebook to produce publication-ready versions of figures and other artifacts to be used in manuscripts. Make sure you include your name as well as contact information for yourself and a future contact in your lab that can answer basic questions about the code. Documenting the beginning and end date of your analysis is also a good idea and can highlight the effort that you have put into the development of the notebook.

## Rule 3: Use cell divisions to make steps clear

Notebooks are an interactive environment, so it is very easy to write and run one-line cells. This supports experimentation but can leave your notebooks messy and full of short fragments that are hard to follow. Instead, try to make each cell in your notebook perform one meaningful step of the analysis that is easy to understand from the code in the cell or the surrounding markdown description. Modularize your code by cells and label the cells with markdown above the cell. Think of each cell as being one paragraph, having one function, or accomplishing one task (for example, create a plot).

Avoid long cells (we suggest that anything over 100 lines or one page is too long). Put low-level documentation in code comments. Use descriptive markdown headers to organize your notebook into sections that can be used to easily navigate the notebook and add a table of contents. Split long notebooks into a series of notebooks and keep a top-level index notebook with links to the individual notebooks. Using clear cell and notebook divisions will make your analysis much easier to read.

## Rule 4: Modularize code

It is always good practice to avoid duplicate code, but in notebooks, it is especially easy to copy a cell, tweak a few lines, paste the resulting code into a new cell or another notebook, and run it again. This form of experimentation is expedient but makes notebooks difficult to read and nearly impossible to maintain if you want to change the functionality of or fix a bug in the copied code. Instead, wrap code you are about to copy and reuse in a function, which you can then call from as many cells as desired. If you are going to reuse the code in other projects or notebooks, consider turning it into a module, package, or library.

Not only does modularization save space, support maintenance, and ease debugging, it also makes it easier to add interactivity. For example, you can tie widgets (ipywidgets, https://ipywidgets.readthedocs.io/en/stable/) to functions to support exploration of different parameter values or support interaction with visualizations without needing to modify the code. This is one way you can design your notebook to be explored (Rule 9).

## Rule 5: Record dependencies

Rerunning your analysis in the future will require accessing not only your code but also any module or library that your code relied on. As is best practice across computational science, manage your dependencies using a package or environment manager like pip or Conda. These enable you to download modules and libraries, specify the version of each you want to use in your analysis, and even generate files such as Conda's environment.yml or pip's requirements.txt that concisely describe all of your dependencies. These files can be used by tools such as Binder or Docker to generate a "container" that other researchers can use to reproduce your analysis using the same versions of every module and library as you did. Always conduct your work in an environment created only from these dependencies to ensure you do not add undocumented dependencies.

As an extra precaution in notebooks, you can explicitly print out your dependencies using a notebook extension such as watermark (https://github.com/rasbt/watermark). Listing the versions of critical dependencies in the notebook itself (best done at the bottom) will ensure that, if used in isolation from its environment, the notebook still contains critical information to help readers run it.

## Rule 6: Use version control

Version control is a critical adjunct to notebook use because the interactive nature of notebooks makes it easy to accidentally change or delete important content. Furthermore, since notebooks contain code and code inevitably contains bugs, being able to determine the history of when a given bug you have discovered was introduced to the code versus when it was fixed—and thus what analyses it may have affected—is a key capability in scientific computation. Consult the Ten Simple Rules paper by Perez-Riverol and colleagues [15] on how to take advantage of Git and GitHub for version control generally. Also follow best practices for organizing your repository for easy version control, for example, http://drivendata.github.io/cookiecutter-data-science/.

However, be aware that Jupyter Notebooks store both code and extensive metadata about each cell as a text file in the JavaScript Object Notation (JSON) format. Version control systems compare differences in these JSON files, not differences in the user-friendly notebook graphical user interface (GUI). Because of this, reported differences between versions of a given notebook are usually difficult for users to find and understand because they are expressed as changes in the abstruse JSON metadata for the notebook. One way to address this issue is to use a notebook-specific diffing tool like nbdime that understands notebook structure and presents differences in meaningful ways (https://github.com/jupyter/nbdime). Another approach is to convert your notebook to a more version-control–friendly filetype such as .py before committing changes.

## Rule 7: Build a pipeline

Notebooks documenting initial, exploratory investigations will rarely be widely generalizable, but once a stable analysis approach has been identified, a well-designed notebook can be generalized into a pipeline that easily repeats that analysis using different input data and parameters. With this end in mind, design your notebook from the beginning to allow such future repurposing. Place key variable declarations, especially those that will be changed when doing a new analysis, at the top of the notebook rather than burying them somewhere in the middle. Perform preparatory steps, like data cleaning, directly in the notebook and avoid manual interventions.

Because notebooks' interactivity make them vulnerable to accidental overwriting or deletion of critical steps by the user, if your analysis runs quickly, make a habit of regularly restarting your kernel and rerunning all cells to make sure you did not accidentally delete a step while cleaning your notebook (and if you did, retrieve the code for it from version control). Restarting your kernel and running all cells is also a good final test of results. To allow partial execution of complex analyses, break long notebooks into smaller notebooks that focus on one or a few analysis steps. Then, ensure that each notebook stores serialized versions of key intermediate results to disk for subsequent notebooks to use.

Once a notebook has been developed, it can be parameterized with a tool such as papermill (https://github.com/nteract/papermill). Such notebooks can be used not only interactively but also as command-line tools that can be executed automatically—a great boon for pipelines! Consider linking your analysis pipeline steps via a Makefile or similar tool that allows for complete noninteractive execution of the entire pipeline, either in full or partial steps. Such automation also supports code quality techniques like software testing; consider testing your workflows from end to end each time a change is committed by integrating your repository to a Continuous Integration system (for example, https://travis-ci.org/). Last but not least, be aware that pipeline notebooks will almost certainly have a very different story (Rule 1) than the initial analyses that engendered them! Remember to remove any introduction, interpretation, or conclusion text that is not universally applicable to different inputs and results and instead replace it with guidance for the pipeline user on how to run and interpret its (potentially novel) results.

## Rule 8: Share and explain your data

Having access to a clearly annotated notebook is of little use to those wanting to reproduce or extend your results if the underlying data are locked away. Strive to make your data or a sample of your data publicly available along with the notebook. While sharing your data takes careful planning, notebooks make it easy to provide a description of your input data and upstream processing steps, which are essential for interpreting results.

Ideally, you will share your entire data set alongside your notebooks. We realize many data sets are too large or too sensitive to share this way. In these cases, consider breaking down large and complex data sets into tiers such that, even if the raw data are prohibitively large to include alongside your published notebooks or are constrained by privacy or other access issues, reproducibility and interpretability isn't lost. You can host public copies of medium-sized, anonymized data in a variety of hosting services (for example, figshare [https://figshare.com/], zenodo [https://zenodo.org/]), and include further processed data sets alongside the notebooks in the final repository. To uniquely and permanently identify data sets, these hosting services provide Digital Object Identifiers (dois). This tiered approach both provides public confidence and allows others to reproduce and reuse the latter stages of an analysis even without access to the full, raw data set.

## Rule 9: Design your notebooks to be read, run, and explored

If you have followed the previous rules, your notebooks should capture your entire process and be easy to read. But how will others access, run, and explore them? There are a number of ways you can support others' reuse of your notebooks. First, store your notebooks in a public code repository with a clear README file and a liberal open source license (https://opensource.org/licenses) granting permission to reuse your code.

Read: Beyond granting permission to reuse your notebook, consider how you can leverage the unique structure of notebooks to support reading. At the very least, leave static HTML/PDF versions of all notebooks stored in the final version of the repository accompanying a publication. If, in 20 years, all other execution technology fails, these are likely to still provide a readable archival record, and with a full dependences list, future users are more likely to be able to recreate the compute environment. You can also use Nbviewer (https://nbviewer.jupyter.org/) to provide static views of your executed notebook online without needing to convert it to a PDF/HTML document first. GitHub uses this service to render any notebooks on their site, so pushing a notebook to GitHub is another good way to make static views easily available. In both cases, you can point collaborators to a URL where they can read through your notebook online.

Run: To support others running your notebooks, you can use Binder [16] to provide a zero-install environment to run your notebooks in the cloud (https://mybinder.org/). Binder enables community members to rerun your notebook online without needing to install Jupyter Notebook or Jupyter Lab on their own machine. More generally, you can create a portable containerized environment, for example, a Docker image (https://docs.docker.com/), or create a dependency description file (see Rule 3) so future users of your notebook can more easily replicate your computing environment when rerunning your notebook.

Explore: Beyond simply replicating the analysis in your notebook, consider how you can design your notebook so future users can tweak and explore your analysis. Consider using ipywidgets (https://ipywidgets.readthedocs.io/en/stable/) to enable future users to change parameters using graphical elements such as dropdowns and sliders rather than tweaking code. Beyond enabling future users to change parameters or insert their own data set, consider how they might want remix or reuse portions of your notebook (perhaps only the data cleaning or plotting steps) and use cell-structure and functions to make it easier to extract these sections (Rule 7).

## Rule 10: Advocate for open research

Clearly, the mere use of a computational notebook does not guarantee others will be able to read, run, or explore your analysis. If the convenience and interactivity of this technology has convinced you to adopt it, take the next step and become an advocate in your lab or workplace in promoting its effective use. Ask lab-mates or colleagues to try to run one of your notebooks and then listen when they explain any difficulties. Try to run their notebooks and let them know if you hit snags. Commit yourself to robust and reproducible analyses as key element of all your research group's computational work, not a phase performed after an analysis is complete or an afterthought triggered by journal or reviewer demands.

## Annotated notebooks

To demonstrate the 10 rules, we have created a Git Repository with annotated example notebooks (https://github.com/jupyter-guide/ten-rules-jupyter). Following Rule 9, read, run, and explore these notebooks. In addition, we have created a repository (https://github.com/jupyter-guide/jupyter-guide) to crowdsource more technical and in-depth tutorials and to keep up with the rapidly evolving Jupyter ecosystem. We encourage you to contribute and share your experiences and know-how following Rule 10.

## Conclusions

Robust and reproducible analyses lie at the heart of science, and several papers have already provided excellent general advice for how to perform and document computational science. However, the advent of computational notebooks presents new opportunities and challenges, both easing precise documentation of complex workflows, and complicating it by means of interactivity. We present 10 simple rules for writing and sharing analyses in Jupyter Notebooks, focusing on annotation of the analysis, organization of code, and ease of access and reuse. Informed by our experience, we hope they contribute to the ecosystem of individuals, labs, publishers, and organizations using notebooks to perform and share computational research.

## Acknowledgments

This paper represents a summary of the workshop "Reproducible Research and Interactive Education—Application of Jupyter Notebooks" held at UC San Diego on April 5, 2018. We thank all participants who contributed ideas to this paper, including Tiago Leao, Nathan Mih, Shweta Purawat, Michael Reich, Britton Smith, Shuai Tang, and Guorong Xu.

## References

1. Barba LA. The hard road to reproducibility. Science. 2016; 354: 142. https://doi.org/10.1126/science.354.6308.142 PMID: 27846503
2. Peng RD. Reproducible Research in Computational Science. Science. 2011; 334: 1226–1227. https://doi.org/10.1126/science.1213847 PMID: 22144613
3. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good enough practices in scientific computing. PLoS Comput Biol. 2017; 13(6):e1005510. https://doi.org/10.1371/journal.pcbi.1005510 PMID: 28640806
4. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comput Biol. 2013; 9(10):e1003285. https://doi.org/10.1371/journal.pcbi.1003285 PMID: 24204232
5. Taschuk M and Wilson G. Ten simple rules for making research software more robust. PLoS Comput. Biol. 2017; 13(4):e1005412. https://doi.org/10.1371/journal.pcbi.1005412 PMID: 28407023
6. Reproducible Research using Jupyter Notebooks. [Internet] [cited 4 Oct 2018]. Available from: https://reproducible-science-curriculum.github.io/workshop-RR-Jupyter/.

7. Pérez F, Granger BE. Computational Narratives as the Engine of Collaborative Data Science. 2015. [Internet] [cited 4 Oct 2018]. Available from: https://blog.jupyter.org/project-jupyter-computational-narratives-as-the-engine-of-collaborative-data-science-2b5fb94c3c58.

8. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B, editors. Positioning and Power in Academic Publishing: Players, Agents and Agendas. Amsterdam: IOS Press; 2016. p. 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

9. Knuth DE. Literate programming. The Computer Journal. 1984; 27(2):97–111.

10. Estimate of Public Jupyter Notebooks on GitHub. [Internet] [cited 4 Oct 2018]. Available from: https://github.com/parente/nbestimate.

11. Rule A, Tabard A, Hollan JD. Exploration and Explanation in Computational Notebooks. CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. New York: ACM; 2018. https://doi.org/10.1145/3173574.3173606

12. Grus, J. I Don't Like Notebooks. JupyterCon. New York, NY. 2018. [Internet] [cited 3 Jan 2019]. Available from: https://docs.google.com/presentation/d/1n2RlMdmv1p25Xy5thJUhkKGvjtV-dkAIsUXP-AL4ffI/edit#slide=id.g3d168d2fd3_0_255

13. Woodbridge M, Sanz D, Mietchen D, Mounce R. Jupyter Notebooks and reproducible data science. 2017. [Internet] [cited 4 Oct 2018]. Available from: https://markwoodbridge.com/2017/03/05/jupyter-reproducible-science.html.

14. Kery MB, Radensky M, Arya M, John BE, Myers BA. The Story in the Notebook: Exploratory Data Science using a Literate Programming Tool. CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. New York: ACM; 2018. https://doi.org/10.1145/3173574.3173748

15. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, et al. Ten Simple Rules for Taking Advantage of Git and GitHub. PLoS Comput. Biol. 2016; 12(7):e1004947. https://doi.org/10.1371/journal.pcbi.1004947 PMID: 27415786

16. Project Jupyter, Bussonnier M, Forde J, Freeman J, Granger B, et al. Binder 2.0—Reproducible, interactive, shareable environments for science at scale. Proceedings of the 17th Python in Science Conference 2018. 2018. p. 113–120. https://doi.org/10.25080/Majora-4af1f417-011

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for carrying out and writing meta-analyses

**Diego A. Forero[1,2]*, Sandra Lopez-Leon[3], Yeimy González-Giraldo[4], Pantelis G. Bagos[5]**

**1** Laboratory of NeuroPsychiatric Genetics, Biomedical Sciences Research Group, School of Medicine, Universidad Antonio Nariño, Bogotá, Colombia, **2** PhD Program in Health Sciences, School of Medicine, Universidad Antonio Nariño, Bogotá, Colombia, **3** Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, United States of America, **4** Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá., Colombia, **5** Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece

* diego.forero@uan.edu.co.



OPEN ACCESS

**Citation:** Forero DA, Lopez-Leon S, González-Giraldo Y, Bagos PG (2019) Ten simple rules for carrying out and writing meta-analyses. PLoS Comput Biol 15(5): e1006922. https://doi.org/10.1371/journal.pcbi.1006922

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** May 16, 2019



**Funding:** YG-G is supported by a PhD fellowship from Centro de Estudios Interdisciplinarios Básicos y Aplicados CEIBA (Rodolfo Llinás Program). DAF is supported by research grants from Colciencias and VCTI. PGB is partially supported by ELIXIR-GR, the Greek Research Infrastructure for data management and analysis in the biosciences. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

In the context of evidence-based medicine, meta-analyses provide novel and useful information [1], as they are at the top of the pyramid of evidence and consolidate previous evidence published in multiple previous reports [2]. Meta-analysis is a powerful tool to cumulate and summarize the knowledge in a research field [3]. Because of the significant increase in the published scientific literature in recent years, there has also been an important growth in the number of meta-analyses for a large number of topics [4]. It has been found that meta-analyses are among the types of publications that usually receive a larger number of citations in the biomedical sciences [5,6]. The methods and standards for carrying out meta-analyses have evolved in recent years [7–9].

Although there are several published articles describing comprehensive guidelines for specific types of meta-analyses, there is still the need for an abridged article with general and updated recommendations for researchers interested in the development of meta-analyses. We present here ten simple rules for carrying out and writing meta-analyses.

## Rule 1: Specify the topic and type of the meta-analysis

Considering that a systematic review [10] is fundamental for a meta-analysis, you can use the Population, Intervention, Comparison, Outcome (PICO) model to formulate the research question. It is important to verify that there are no published meta-analyses on the specific topic in order to avoid duplication of efforts [11]. In some cases, an updated meta-analysis in a topic is needed if additional data become available. It is possible to carry out meta-analyses for multiple types of studies, such as epidemiological variables for case-control, cohort, and randomized clinical trials. As observational studies have a larger possibility of having several biases, meta-analyses of these types of designs should take that into account. In addition, there is the possibility to carry out meta-analyses for genetic association studies, gene expression studies, genome-wide association studies (GWASs), or data from animal experiments. It is advisable to preregister the systematic review protocols at the International Prospective Register of Systematic Reviews (PROSPERO; https://www.crd.york.ac.uk/Prospero) database [12]. Keep in mind that an increasing number of journals require registration prior to publication.

## Rule 2: Follow available guidelines for different types of meta-analyses

There are several available general guidelines. The first of such efforts were the Quality of Reports of Meta-analyses of Randomized Controlled Trials (QUORUM) [13] and the Meta-analysis of Observational Studies in Epidemiology (MOOSE) statements [14], but currently, the Preferred Reporting Items for Systematic reviews and Meta-analyses (PRISMA) [15] has been broadly cited and used. In addition, there have been efforts to develop specific guidelines regarding meta-analyses for clinical studies (Cochrane Handbook; https://training.cochrane.org/handbook), genetic association studies [16], genome-wide expression studies [17], GWASs [18], and animal studies [19].

## Rule 3: Establish inclusion criteria and define key variables

You should establish in advance the inclusion (such as type of study, language of publication, among others) and exclusion (such as minimal sample size, among others) criteria. Keep in mind that the current consensus advises against strict criteria concerning language or sample size. You should clearly define the variables that will be extracted from each primary article. Broad inclusion criteria increase heterogeneity between studies, and narrow inclusion criteria can make it difficult to find studies; therefore, a compromise should be found. Prospective meta-analyses, which usually are carried out by international consortia, have the advantage of the possibility of including individual-level data [20].

## Rule 4: Carry out a systematic search in different databases and extract key data

You can carry out your systematic search in several bibliographic databases, such as PubMed, Embase, The Cochrane Central Register of Controlled Trials, Scopus, Web of Science, and Google Scholar [21]. Usually, searching in several databases helps to minimize the possibility of failing to identify all published studies [22]. In some specific areas, searching in specialized databases is also worth doing (such as BIOSIS, Cumulative index to Nursing and Allied Health Literature (CINAHL), PsycINFO, Sociological Abstracts, and EconLit, among others). Moreover, in other cases, direct search for the data is also advisable (i.e., Gene Expression Omnibus [GEO] database for gene expression studies) [23]. Usually, the bibliography of review articles might help to identify additional articles and data from other types of documents (such as theses or conference proceedings) that might be included in your meta-analysis. The Web of Science database can be used to identify publications that have cited key articles. Adequate extraction and recording of key data from primary articles are fundamental for carrying out a meta-analysis. Quality assessment of the included studies is also an important issue; it can be used for determining inclusion criteria, sensitivity analysis, or differential weighting of the studies. For example the Jadad scale [24] is frequently used for randomized clinical trials, the Newcastle–Ottawa scale [25] for nonrandomized studies, and QUADAS-2 for the Quality Assessment of Diagnostic Accuracy Studies [26]. It is recommended that these steps be carried out by two researchers in parallel and that discrepancies be resolved by consensus. Nevertheless, the reader must be aware that quality assessment has been criticized, especially when it reduces the studies to a single "quality" score [27,28]. In any case, it is important to avoid the confusion of using guidelines for the reporting of primary studies as scales for the assessment of the quality of included articles [29,30].

## Rule 5: Contact authors of primary articles to ask for missing data

It is common that key data are not available in the main text or supplementary files of primary articles [31], leading to the need to contact the authors to ask for missing data. However, the rate of response from authors is lower than expected. There are multiple standards that promote the availability of primary data in published articles, such as the minimum information about a microarray experiment (MIAME) [32] and the STrengthening the REporting of

Genetic Association Studies (STREGA) [33]. In some areas, such as genetics, in which it was shown that it is possible to identify an individual using the aggregated statistics from a particular study [34], strict criteria are imposed for data sharing, and specialized permissions might be needed.

## Rule 6: Select the best statistical models for your question

For cases in which there is enough primary data of adequate quality for a quantitative summary, there is the option to carry out a meta-analysis. The potential analyst must be warned that in many cases the data are reported in noncompatible forms, so one must be ready to perform various types of transformations. Thankfully, there are methods available for extracting and transforming data regarding continuous variables [35–37], $2 \times 2$ tables [38,39], or survival data [40]. Frequently, meta-analyses are based on fixed-effects or random-effects statistical models [20]. In addition, models based on combining ranks or *p*-values are also available and can be used in specific cases [41–44]. For more complex data, multivariate methods for meta-analysis have been proposed [45,46]. Additional statistical examinations involve sensitivity analyses, metaregressions, subgroup analyses, and calculation of heterogeneity metrics, such as Q or $I^2$ [20]. It is fundamental to assess and, if present, explain the possible sources of heterogeneity. Although random-effects models are suitable for cases of between-studies heterogeneity, the sources of between-studies variation should be identified, and their impact on effect size should be quantified using statistical tests, such as subgroup analyses or metaregression. Publication bias is an important aspect to consider [47], since in many cases negative findings have less probability of being published. Other types of bias, such as the so-called "Proteus phenomenon" [48] or "winner's curse" [49], are common in some scientific fields, such as genetics, and the approach of cumulative meta-analysis is suggested in order to identify them.

## Rule 7: Use available software to carry metastatistics

There are several very user-friendly and freely available programs for carrying out meta-analyses [43,44], either within the framework of a statistical package such as Stata or R or as stand-alone applications. Stata and R [50–52] have dozens of routines, mostly user written, that can handle most meta-analysis tasks, even complex analyses such as network meta-analysis and meta-analyses of GWASs and gene expression studies (https://cran.r-project.org/web/views/MetaAnalysis.html; https://www.stata.com/support/faqs/statistics/meta-analysis). There are also stand-alone packages that can be useful for general applications or for specific areas, such as OpenMetaAnalyst [53], NetworkAnalyst [54], JASP [55], MetaGenyo [56], Cochrane RevMan (https://community.cochrane.org/help/tools-and-software/revman-5), EpiSheet (krothman.org/episheet.xls), GWAR [57], GWAMA [58], and METAL [59]. Some of these programs are web services or stand-alone software. In some cases, certain programs can present issues when they are run because of their dependency on other packages.

## Rule 8: The records and study report must be complete and transparent

Following published guidelines for meta-analyses guarantees that the manuscript will describe the different steps and methods used, facilitating their transparency and replicability [15]. Data such as search and inclusion criteria, numbers of abstracts screened, and included studies are quite useful, in addition to details of meta-analytical strategies used. An assessment of quality of included studies is also useful [60]. A spreadsheet can be constructed in which every step in the selection criteria is recorded; this will be helpful to construct flow charts. In this context, a flow diagram describing the progression between the different steps is quite useful and might enhance the quality of the meta-analysis [61]. Records will be also useful if, in the future, the meta-analysis needs to be updated. Stating the limitations of the analysis is also important [62].

## Rule 9: Provide enough data in your manuscript

A table with complete information about included studies (such as author, year, details of included subjects, DOIs, or PubMed IDs, among others) is quite useful in an article reporting a meta-analysis; it can be included in the main text of the manuscript or as a supplementary file. Software used for carrying out meta-analyses and to generate key graphs, such as forest plots, should be referenced. Summary effect measures, such as a pooled odds ratios or the counts used to generate them, should be always reported, including confidence intervals. It is also possible to generate figures with information from multiple forest plots [63]. In the case of positive findings, plots from sensitivity analyses are quite informative. In more-complex analyses, it is advisable to include in the supplementary files the scripts used to generate the results [64].

## Rule 10: Provide context for your findings and suggest future directions

The Discussion section is an important scientific component in a manuscript describing a meta-analysis, as the authors should discuss their current findings in the context of the available scientific literature and existing knowledge [65]. Authors can discuss possible reasons for the positive or negative results of their meta-analysis, provide an interpretation of findings based on available biological or epidemiological evidence, and comment on particular features of individual studies or experimental designs used [66]. As meta-analyses are usually synthesizing the existing evidence from multiple primary studies, which commonly took years and large amounts of funding, authors can recommend key suggestions for conducting and/or reporting future primary studies [67].

As open science is becoming more important around the globe [68,69], adherence to published standards, in addition to the evolution of methods for different meta-analytical applications, will be even more important to carry out meta-analyses of high quality and impact.

## References


1. Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, et al. (2014) How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA 312: 171–179. https://doi.org/10.1001/jama.2014.5559 PMID: 25005654
2. Garg AX, Hackam D, Tonelli M (2008) Systematic review and meta-analysis: when one study is just not enough. Clin J Am Soc Nephrol 3: 253–260. https://doi.org/10.2215/CJN.01430307 PMID: 18178786
3. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G (2013) Meta-analysis: pitfalls and hints. Heart Lung Vessel 5: 219–225. PMID: 24364016
4. Ioannidis JP, Chang CQ, Lam TK, Schully SD, Khoury MJ (2013) The geometric increase in meta-analyses from China in the genomic era. PLoS ONE 8: e65602. https://doi.org/10.1371/journal.pone.0065602 PMID: 23776510
5. Uthman OA, Okwundu CI, Wiysonge CS, Young T, Clarke A (2013) Citation classics in systematic reviews and meta-analyses: who wrote the top 100 most cited articles? PLoS ONE 8: e78517. https://doi.org/10.1371/journal.pone.0078517 PMID: 24155987
6. Patsopoulos NA, Analatos AA, Ioannidis JP (2005) Relative citation impact of various study designs in the health sciences. JAMA 293: 2362–2366. https://doi.org/10.1001/jama.293.19.2362 PMID: 15900006
7. Sutton AJ, Higgins JP (2008) Recent developments in meta-analysis. Stat Med 27: 625–650. https://doi.org/10.1002/sim.2934 PMID: 17590884
8. Hedges LV (2015) The early history of meta-analysis. Res Synth Methods 6: 284–286. https://doi.org/10.1002/jrsm.1149 PMID: 26097046

9. Glass GV (2015) Meta-analysis at middle age: a personal history. Res Synth Methods 6: 221–231. https://doi.org/10.1002/jrsm.1133 PMID: 26355796
10. Pautasso M (2013) Ten simple rules for writing a literature review. PLoS Comput Biol 9: e1003149. https://doi.org/10.1371/journal.pcbi.1003149 PMID: 23874189
11. Siontis KC, Hernandez-Boussard T, Ioannidis JP (2013) Overlapping meta-analyses on the same topic: survey of published studies. BMJ 347: f4501. https://doi.org/10.1136/bmj.f4501 PMID: 23873947
12. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, et al. (2013) PROSPERO at one year: an evaluation of its utility. Syst Rev 2: 4. https://doi.org/10.1186/2046-4053-2-4 PMID: 23320413
13. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, et al. (2000) Improving the Quality of Reports of Meta-Analyses of Randomised Controlled Trials: The QUOROM Statement. Onkologie 23: 597–602. https://doi.org/10.1159/000055014 PMID: 11441269
14. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, et al. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 283: 2008–2012. PMID: 10789670
15. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6: e1000097. https://doi.org/10.1371/journal.pmed.1000097 PMID: 19621072
16. Sagoo GS, Little J, Higgins JP (2009) Systematic reviews of genetic association studies. Human Genome Epidemiology Network. PLoS Med 6: e28. https://doi.org/10.1371/journal.pmed.1000028 PMID: 19260758
17. Ramasamy A, Mondry A, Holmes CC, Altman DG (2008) Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Med 5: e184. https://doi.org/10.1371/journal.pmed.0050184 PMID: 18767902
18. Evangelou E, Ioannidis JP (2013) Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14: 379–389. https://doi.org/10.1038/nrg3472 PMID: 23657481
19. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, et al. (2014) Meta-analysis of data from animal studies: a practical guide. J Neurosci Methods 221: 92–102. https://doi.org/10.1016/j.jneumeth.2013.09.010 PMID: 24099992
20. Kavvoura FK, Ioannidis JP (2008) Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet 123: 1–14. https://doi.org/10.1007/s00439-007-0445-9 PMID: 18026754
21. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G (2008) Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 22: 338–342. https://doi.org/10.1096/fj.07-9492LSF PMID: 17884971
22. Lemeshow AR, Blum RE, Berlin JA, Stoto MA, Colditz GA (2005) Searching one or two databases was insufficient for meta-analysis of observational studies. J Clin Epidemiol 58: 867–873. https://doi.org/10.1016/j.jclinepi.2005.03.004 PMID: 16085190
23. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2011) NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic Acids Res 39: D1005–1010. https://doi.org/10.1093/nar/gkq1184 PMID: 21097893
24. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, et al. (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 17: 1–12. PMID: 8721797
25. Stang A (2010) Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol 25: 603–605. https://doi.org/10.1007/s10654-010-9491-z PMID: 20652370
26. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, et al. (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155: 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009 PMID: 22007046
27. Greenland S, O'Rourke K (2001) On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics 2: 463–471. https://doi.org/10.1093/biostatistics/2.4.463 PMID: 12933636
28. Juni P, Witschi A, Bloch R, Egger M (1999) The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 282: 1054–1060. PMID: 10493204
29. da Costa BR, Cevallos M, Altman DG, Rutjes AW, Egger M (2011) Uses and misuses of the STROBE statement: bibliographic study. BMJ Open 1: e000048. https://doi.org/10.1136/bmjopen-2010-000048 PMID: 22021739

30. Harrison JK, Reid J, Quinn TJ, Shenkin SD (2017) Using quality assessment tools to critically appraise ageing research: a guide for clinicians. Age Ageing 46: 359–365. https://doi.org/10.1093/ageing/afw223 PMID: 27932357

31. Ioannidis JP, Allison DB, Ball CA, Coulibaly I, Cui X, et al. (2009) Repeatability of published microarray gene expression analyses. Nat Genet 41: 149–155. https://doi.org/10.1038/ng.295 PMID: 19174838

32. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29: 365–371. https://doi.org/10.1038/ng1201-365 PMID: 11726920

33. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, et al. (2009) STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 6: e22. https://doi.org/10.1371/journal.pmed.1000022 PMID: 19192942

34. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, et al. (2008) Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet 4: e1000167. https://doi.org/10.1371/journal.pgen.1000167 PMID: 18769715

35. Chene G, Thompson SG (1996) Methods for summarizing the risk associations of quantitative variables in epidemiologic studies in a consistent form. Am J Epidemiol 144: 610–621. PMID: 8797521

36. Hozo SP, Djulbegovic B, Hozo I (2005) Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 5: 13. https://doi.org/10.1186/1471-2288-5-13 PMID: 15840177

37. da Costa BR, Rutjes AW, Johnston BC, Reichenbach S, Nuesch E, et al. (2012) Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. Int J Epidemiol 41: 1445–1459. https://doi.org/10.1093/ije/dys124 PMID: 23045205

38. Di Pietrantonj C (2006) Four-fold table cell frequencies imputation in meta analysis. Stat Med 25: 2299–2322. https://doi.org/10.1002/sim.2287 PMID: 16025540

39. Hirji KF, Fagerland MW (2011) Calculating unreported confidence intervals for paired data. BMC Med Res Methodol 11: 66. https://doi.org/10.1186/1471-2288-11-66 PMID: 21569392

40. Parmar MK, Torri V, Stewart L (1998) Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 17: 2815–2834. PMID: 9921604

41. Begum F, Ghosh D, Tseng GC, Feingold E (2012) Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Res 40: 3777–3784. https://doi.org/10.1093/nar/gkr1255 PMID: 22241776

42. Tseng GC, Ghosh D, Feingold E (2012) Comprehensive literature review and statistical considerations for microarray meta-analysis. Nucleic Acids Res 40: 3785–3799. https://doi.org/10.1093/nar/gkr1265 PMID: 22262733

43. Dimou NL, Pantavou KG, Braliou GG, Bagos PG (2018) Multivariate Methods for Meta-Analysis of Genetic Association Studies. Methods Mol Biol 1793: 157–182. https://doi.org/10.1007/978-1-4939-7868-7_11 PMID: 29876897

44. Kontou PI, Pavlopoulou A, Bagos PG (2018) Methods of Analysis and Meta-Analysis for Identifying Differentially Expressed Genes. Methods Mol Biol 1793: 183–210. https://doi.org/10.1007/978-1-4939-7868-7_12 PMID: 29876898

45. Mavridis D, Salanti G (2013) A practical introduction to multivariate meta-analysis. Stat Methods Med Res 22: 133–158. https://doi.org/10.1177/0962280211432219 PMID: 22275379

46. Jackson D, Riley R, White IR (2011) Multivariate meta-analysis: potential and promise. Stat Med 30: 2481–2498. https://doi.org/10.1002/sim.4172 PMID: 21268052

47. Rothstein HR, Sutton AJ, Borenstein M (2006) Publication bias in meta-analysis: Prevention, assessment and adjustments. Hoboken, NJ: John Wiley & Sons.

48. Ioannidis JP, Trikalinos TA (2005) Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. J Clin Epidemiol 58: 543–549. https://doi.org/10.1016/j.jclinepi.2004.10.019 PMID: 15878467

49. Kraft P (2008) Curses—winner's and otherwise—in genetic epidemiology. Epidemiology 19: 649–651; discussion 657–648. https://doi.org/10.1097/EDE.0b013e318181b865 PMID: 18703928

50. Sterne JA, Bradburn MJ, Egger M (2001) Meta–Analysis in Stata™. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. Hoboken, NJ: Wiley. pp. 347–369.

51. Quintana DS (2015) From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 6: 1549. https://doi.org/10.3389/fpsyg.2015.01549 PMID: 26500598

52. Polanin JR, Hennessy EA, Tanner-Smith EE (2017) A review of meta-analysis packages in R. Journal of Educational and Behavioral Statistics 42: 206–242.

53. Wallace BC, Schmid CH, Lau J, Trikalinos TA (2009) Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. BMC Med Res Methodol 9: 80. https://doi.org/10.1186/1471-2288-9-80 PMID: 19961608

54. Xia J, Benner MJ, Hancock RE (2014) NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res 42: W167–174. https://doi.org/10.1093/nar/gku443 PMID: 24861621

55. Quintana DS, Williams DR (2018) Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. BMC Psychiatry 18: 178. https://doi.org/10.1186/s12888-018-1761-4 PMID: 29879931

56. Martorell-Marugan J, Toro-Dominguez D, Alarcon-Riquelme ME, Carmona-Saez P (2017) MetaGenyo: a web tool for meta-analysis of genetic association studies. BMC Bioinformatics 18: 563. https://doi.org/10.1186/s12859-017-1990-4 PMID: 29246109

57. Dimou NL, Tsirigos KD, Elofsson A, Bagos PG (2017) GWAR: robust analysis and meta-analysis of genome-wide association studies. Bioinformatics 33: 1521–1527. https://doi.org/10.1093/bioinformatics/btx008 PMID: 28108451

58. Magi R, Morris AP (2010) GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11: 288. https://doi.org/10.1186/1471-2105-11-288 PMID: 20509871

59. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190–2191. https://doi.org/10.1093/bioinformatics/btq340 PMID: 20616382

60. Ioannidis JP, Boffetta P, Little J, O'Brien TR, Uitterlinden AG, et al. (2008) Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol 37: 120–132. https://doi.org/10.1093/ije/dym159 PMID: 17898028

61. Vu-Ngoc H, Elawady SS, Mehyar GM, Abdelhamid AH, Mattar OM, et al. (2018) Quality of flow diagram in systematic review and/or meta-analysis. PLoS ONE 13: e0195955. https://doi.org/10.1371/journal.pone.0195955 PMID: 29949595

62. Ioannidis JP (2007) Limitations are not properly acknowledged in the scientific literature. J Clin Epidemiol 60: 324–329. https://doi.org/10.1016/j.jclinepi.2006.09.011 PMID: 17346604

63. Neyeloff JL, Fuchs SC, Moreira LB (2012) Meta-analyses and Forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis. BMC Res Notes 5: 52. https://doi.org/10.1186/1756-0500-5-52 PMID: 22264277

64. Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, et al. (2014) Ten simple rules for effective computational research. PLoS Comput Biol 10: e1003506. https://doi.org/10.1371/journal.pcbi.1003506 PMID: 24675742

65. Russo MW (2007) How to Review a Meta-analysis. Gastroenterol Hepatol (N Y) 3: 637–642.

66. Khan KS, Kunz R, Kleijnen J, Antes G (2003) Five steps to conducting a systematic review. J R Soc Med 96: 118–121. PMID: 12612111

67. Zhang W (2014) Ten simple rules for writing research papers. PLoS Comput Biol 10: e1003453. https://doi.org/10.1371/journal.pcbi.1003453 PMID: 24499936

68. Wallach JD, Boyack KW, Ioannidis JPA (2018) Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. PLoS Biol 16: e2006930. https://doi.org/10.1371/journal.pbio.2006930 PMID: 30457984

69. Masum H, Rao A, Good BM, Todd MH, Edwards AM, et al. (2013) Ten simple rules for cultivating open science and collaborative R&D. PLoS Comput Biol 9: e1003244. https://doi.org/10.1371/journal.pcbi.1003244 PMID: 24086123

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules on how to create open access and reproducible molecular simulations of biological systems

**Arne Elofsson[1]*, Berk Hess[2], Erik Lindahl[1,2], Alexey Onufriev[3], David van der Spoel[4], Anders Wallqvist[5]**

1 Department of Biochemistry and Biophysics, Science for life Laboratory, Stockholm University, Solna, Sweden, 2 Department of Physics, Swedish e-Science Research Center, KTH Royal Institute of Technology, Stockholm, Sweden, 3 Departments of Computer Science and Physics, Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia, United States of America, 4 Uppsala Centre for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden, 5 Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland, United States of America

* arne@bioinfo.se



OPEN ACCESS

**Citation:** Elofsson A, Hess B, Lindahl E, Onufriev A, van der Spoel D, Wallqvist A (2019) Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. PLoS Comput Biol 15(1): e1006649. https://doi.org/10.1371/journal.pcbi.1006649

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** January 17, 2019



**Funding:** The authors received so specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

All *PLOS* journals have an open data policy that, amongst other things, states that all data and related metadata underlying the findings reported in a submitted manuscript should be deposited in an appropriate public repository, or for smaller datasets, as supporting information. This should obviously apply to computational methods as well, but unfortunately this is not always applied in practice, although it is of greatest importance for the scientific quality of simulations [1] and other modeling projects [2].

Molecular dynamics [3] and other type of simulations [2,4] have become a fundamental part of life sciences. The simulations are dependent on a number of parameters such as force fields, initial configurations, simulation protocols, and software. Researchers have different opinions about the types of software they prefer, and in general, we believe authors should be free to choose the tools that best fit their needs. However, as scientists, we also have a common obligation to critically test each other's statements to find mistakes (including errors in the algorithms and bugs in the code), which can be exemplified by a heated debate over simulations of supercooled water that ended up being due to a subtle algorithmic issue [5], and we believe *PLOS* has a particularly strong responsibility to lead this development even if it might cause some short-term grief [6].

In particular, all published results should, in principle, be possible to reproduce independently by scientists in other labs using different tools. To ensure this, we propose a set of standards that any publication in *PLOS Computational Biology*, and hopefully, publications in other journals as well, should follow. We do believe that the sooner such policies are widely adapted, the more open and collaborative science will flourish [7].

These 10 simple rules should not be limited to molecular dynamics but also include Monte Carlo simulations, quantum mechanics calculations, molecular docking, and any other computational methods involving computations on biological molecules.

## Rule 1: The simulation protocol should be provided

The complete set of input files that are used in the simulations should be provided, either as supplementary material or preferably through a publicly available repository.

## Rule 2: Topology and parameters should be accessible for everyone

All topology and parameter files used in the simulations should be provided and made publicly available so they can be implemented and tested with a different program if necessary. That means the files should either be in human-readable format or the conversion rules should be be publically available.

## Rule 3: Initial coordinate files should be included

All simulations are strongly dependent on the initial conditions [8]. To ensure maximum reproducibility, the authors should provide the input coordinate files for the simulations in the appropriate formats for the software used. The input files to initiate the simulations should be provided in a format ensuring that a reader can repeat all parts of the calculation workflow himself or herself. That means the files should either be in human-readable format or the conversion rules should be publicly available.

## Rule 4: Full information about all software used needs to be provided

Reviewers and readers must be able to reproduce results with as much detail as possible. This means authors need to provide enough details so that the work can be repeated with widely available programs or that the software is provided. In particular, indicate the specific version of the software package used in the simulation. To further improve reproducibility, we encourage software authors to add information about compilers, flags, and the hardware used to log files.

## Rule 5: Simulation results should be deposited in a database

Following the *PLOS* editorial policy for data access, the authors should deposit representative snapshots from the trajectories and/or simulations in findable, accessible, interoperable, reusable (FAIR) public repositories. The deposited snapshots must be dense enough so that the reported biological insights are supported with the same statistical error margin as originally reported and so that new analyses and publications can be performed using the deposited data.

## Rule 6: Results should be easy to reproduce

Although the advantages of open source software are plentiful [9], many authors still use commercial software for simulations. However, in these cases, if the software used is not publicly available, the simulation method must be provided or already published in sufficient details so that the results can be reproduced within reported margin of error using publicly available software. Software and scripts used for analysis must also be made publicly available.

## Rule 7: In docking studies, details should be included

For all studies including screening and docking, the complete set of molecules tested as well as the scoring functions used and the high-ranking poses should be publicly available either as databases or detailed descriptions.

## Rule 8: In quantum mechanics calculations, all energies should be included in the results

For quantum mechanics studies, the authors need to provide the following information: absolute energies and energy breakdowns, the level of theory used, the basis set used, the optimization algorithm used, and coordinates of all optimized stationary points. Ideally, the archive entry for each calculation will be provided alongside the coordinates.

## Rule 9: All sampling-based results should be evaluated using proper statistics

As in all scientific studies, statistical rigor is necessary in computational studies to evaluate the significance of an observation—in particular, for any method based on sampling, such as molecular dynamics or Monte Carlo simulations. Appropriate estimates of statistical uncertainty are therefore necessary and should be included for each relevant finding.

## Rule 10: Be Nice

Remember that we all are a community, so sharing your data, methodologies, software, and results in such a way so that others can use it will make the entire community thrive. This also applies to the readers—they should not expect unlimited support for using in-house software or methods. Just the fact that it is available provides an important resource for the community.

## Conclusion

As a result of these discussions, *PLOS Computational Biology* has made the following extension to the *PLOS* data sharing policies:

A. The authors should provide a README file with a list of included files and/or links to publicly available repositories along with their brief description.

B. The authors should describe all software used including the specific version(s) used in the work and how it can be obtained.

C. *PLOS* expects researchers to share software and scripts needed for the work. If this cannot be made publicly available (e.g., due to licenses), the simulation method should be provided in sufficient detail so the results can, in principle, be reproduced using publicly available software.

D. The authors should provide the complete set of input files used to initiate the calculations, including input coordinates, topologies, and parameter files. These files must be provided in human-readable formats and should preferably be included as supplementary material.

E. The authors should deposit trajectories in a public repository according to FAIR data principles [10]. Examples of such databases include ModEL [11], Nomad (https://nomad-repository.eu/), and the Dryad repository (https://datadryad.org/).

## Acknowledgments

All authors contributed equally.

## References


1. Benjamin B, Acharya A, Batista VS. "Is the Supporting Information the Venue for Reproducibility and Transparency?" The Journal of Physical Chemistry. B 121 (51): 11425–26. 2017 https://doi.org/10.1021/acs.jpcb.7b11664 PMID: 29281888
2. Walters B, Patrick W. "Modeling, Informatics, and the Quest for Reproducibility." Journal of Chemical Information and Modeling 53 (7): 1529–30. 2013 https://doi.org/10.1021/ci400197w PMID: 23758509
3. Allen MP, Tildesley DJ. Computer Simulation of Liquids: Second Edition. Oxford University Press. 2017
4. Schlick T. Molecular Modeling and Simulation. 2002
5. Smart AG. "The War over Supercooled Water." Physics Today 22 Aug (2018). https://doi.org/10.1063/pt.6.1.20180822a. 2018
6. Waltemath D, Wolkenhauer O. "How Modeling Standards, Software, and Initiatives Support Reproducibility in Systems Biology and Systems Medicine." IEEE Transactions on Bio-Medical Engineering 63 (10): 1999–2006. 2016 https://doi.org/10.1109/TBME.2016.2555481 PMID: 27295645
7. Masum H, Rao A, Good BM, Todd MH, Edwards AM, Chan L, et al. "Ten Simple Rules for Cultivating Open Science and Collaborative R&D." PLoS Comput Biol 9 (9): e1003244. 2013 https://doi.org/10.1371/journal.pcbi.1003244 PMID: 24086123
8. Zeiske T, Stafford KA, Friesner RA, Palmer AG. "Starting-Structure Dependence of Nanosecond Timescale Intersubstate Transitions and Reproducibility of MD-Derived Order Parameters." Proteins 81 (3): 499–509. 2013 https://doi.org/10.1002/prot.24209 PMID: 23161667
9. Prlić A, Procter JB. "Ten Simple Rules for the Open Development of Scientific Software." PLoS Comput Biol 8 (12): e1002802. 2012 https://doi.org/10.1371/journal.pcbi.1002802 PMID: 23236269
10. Wilkinson MD, Dumontier N, Jsbrand I, Aalbersberg J, Appleton G, Axton M, et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3 (March): 160018. 2016
11. Meyer T, D'Abramo M, Hospital A, Rueda M, Ferrer-Costa C, Pérez A, et al. "MoDEL (Molecular Dynamics Extended Library): A Database of Atomistic Molecular Dynamics Trajectories." Structure 18 (11): 1399–1409. 2010 https://doi.org/10.1016/j.str.2010.07.013 PMID: 21070939

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for responsible big data research

**Matthew Zook[1]*, Solon Barocas[2], danah boyd[2,3], Kate Crawford[2,4], Emily Keller[3], Seeta Peña Gangadharan[5], Alyssa Goodman[6], Rachelle Hollander[7], Barbara A. Koenig[8], Jacob Metcalf[9], Arvind Narayanan[10], Alondra Nelson[11], Frank Pasquale[12]**

**1** Department of Geography, University of Kentucky, Lexington, Kentucky, United States of America, **2** Microsoft Research, New York, New York, United States of America, **3** Data & Society, New York, New York, United States of America, **4** Information Law Institute, New York University, New York, New York, United States of America, **5** Department of Media and Communications, London School of Economics, London, United Kingdom, **6** Harvard-Smithsonian Center for Astrophysics, Harvard University, Cambridge, Massachusetts, United States of America, **7** Center for Engineering Ethics and Society, National Academy of Engineering, Washington, DC, United States of America, **8** Institute for Health Aging, University of California-San Francisco, San Francisco, California, United States of America, **9** Ethical Resolve, Santa Cruz, California, United States of America, **10** Department of Computer Science, Princeton University, Princeton, New Jersey, United States of America, **11** Department of Sociology, Columbia University, New York, New York, United States of America, **12** Carey School of Law, University of Maryland, Baltimore, Maryland, United States of America

* zook@uky.edu



OPEN ACCESS

**Citation:** Zook M, Barocas S, boyd d, Crawford K, Keller E, Gangadharan SP, et al. (2017) Ten simple rules for responsible big data research. PLoS Comput Biol 13(3): e1005399. https://doi.org/10.1371/journal.pcbi.1005399

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** March 30, 2017



**Funding:** The work for this article was supported by the National Science Foundation grant # IIS-1413864. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

The use of big data research methods has grown tremendously over the past five years in both academia and industry. As the size and complexity of available datasets has grown, so too have the ethical questions raised by big data research. These questions become increasingly urgent as data and research agendas move well beyond those typical of the computational and natural sciences, to more directly address sensitive aspects of human behavior, interaction, and health. The tools of big data research are increasingly woven into our daily lives, including mining digital medical records for scientific and economic insights, mapping relationships via social media, capturing individuals' speech and action via sensors, tracking movement across space, shaping police and security policy via "predictive policing," and much more.

The beneficial possibilities for big data in science and industry are tempered by new challenges facing researchers that often lie outside their training and comfort zone. Social scientists now grapple with data structures and cloud computing, while computer scientists must contend with human subject protocols and institutional review boards (IRBs). While the connection between individual datum and actual human beings can appear quite abstract, the scope, scale, and complexity of many forms of big data creates a rich ecosystem in which human participants and their communities are deeply embedded and susceptible to harm. This complexity challenges any normative set of rules and makes devising universal guidelines difficult.

Nevertheless, the need for direction in responsible big data research is evident, and this article provides a set of "ten simple rules" for addressing the complex ethical issues that will inevitably arise. Modeled on *PLOS Computational Biology*'s ongoing collection of rules, the recommendations we outline involve more nuance than the words "simple" and "rules" suggest. This nuance is inevitably tied to our paper's starting premise: all big data research on social, medical, psychological, and economic phenomena engages with human subjects, and researchers have the ethical responsibility to minimize potential harm.

The variety in data sources, research topics, and methodological approaches in big data belies a one-size-fits-all checklist; as a result, these rules are less specific than some might hope. Rather, we exhort researchers to recognize the human participants and complex systems contained within their data and make grappling with ethical questions part of their standard workflow. Towards this end, we structure the first five rules around how to reduce the chance of harm resulting from big data research practices; the second five rules focus on ways researchers can contribute to building best practices that fit their disciplinary and methodological approaches. At the core of these rules, we challenge big data researchers who consider their data disentangled from the ability to harm to reexamine their assumptions. The examples in this paper show how often even seemingly innocuous and anonymized data have produced unanticipated ethical questions and detrimental impacts.

This paper is a result of a two-year National Science Foundation (NSF)-funded project that established the Council for Big Data, Ethics, and Society, a group of 20 scholars from a wide range of social, natural, and computational sciences (http://bdes.datasociety.net/). The Council was charged with providing guidance to the NSF on how to best encourage ethical practices in scientific and engineering research, utilizing big data research methods and infrastructures [1].

## 1. Acknowledge that data are people and can do harm

One of the most fundamental rules of responsible big data research is the steadfast recognition that most data represent or impact people. Simply starting with the assumption that all data are people until proven otherwise places the difficulty of disassociating data from specific individuals front and center. This logic is readily evident for “risky” datasets, e.g., social media with inflammatory language, but even seemingly benign data can contain sensitive and private information, e.g., it is possible to extract data on the exact heart rates of people from YouTube videos [2]. Even data that seemingly have nothing to do with people might impact individuals’ lives in unexpected ways, e.g., oceanographic data that change the risk profiles of communities’ and properties’ values or Exchangeable Image Format (EXIF) records from photos that contain location coordinates and reveal the photographer’s movement or even home location.

Harm can also result when seemingly innocuous datasets about population-wide effects are used to shape the lives of individuals or stigmatize groups, often without procedural recourse [3,4]. For example, social network maps for services such as Twitter can determine credit-worthiness [5], opaque recidivism scores can shape criminal justice decisions in a racially disparate manner [6], and categorization based on zip codes resulted in less access to Amazon Prime same-day delivery service for African-Americans in United States cities [7]. These high-profile cases show that apparently neutral data can yield discriminatory outcomes, thereby compounding social inequities.

Other cases show that “public” datasets are easily adapted for highly invasive research by incorporating other data, such as Hague et al.’s [8] use of property records and geographic profiling techniques to allegedly identify the pseudonymous artist Banksy [9]. In particular, data ungoverned by substantive consent practices, whether social media or the residual DNA we continually leave behind us, may seem public but can cause unintentional breaches of privacy and other harms [9,10].

Start with the assumption that data are people (until proven otherwise), and use it to guide your analysis. No one gets an automatic pass on ethics.

## 2. Recognize that privacy is more than a binary value

Breaches of privacy are key means by which big data research can do harm, and it is important to recognize that privacy is contextual [11] and situational [12], not reducible to a simple public/private binary. Just because something has been shared publicly does not mean any subsequent use would be unproblematic. Looking at a single Instagram photo by an individual has different ethical implications than looking at someone's full history of all social media posts. Privacy depends on the nature of the data, the context in which they were created and obtained, and the expectations and norms of those who are affected. Understand that your attitude towards acceptable use and privacy may not correspond with those whose data you are using, as privacy preferences differ across and within societies.

For example, Tene and Polonetsky [13] explore how pushing past social norms, particularly in novel situations created by new technologies, is perceived by individuals as "creepy" even when they do not violate data protection regulations or privacy laws. Social media apps that utilize users' locations to push information, corporate tracking of individuals' social media and private communications to gain customer intelligence, and marketing based on search patterns have been perceived by some to be "creepy" or even outright breaches of privacy. Likewise, distributing health records is a necessary part of receiving health care, but this same sharing brings new ethical concerns when it goes beyond providers to marketers.

Privacy also goes beyond single individuals and extends to groups [10]. This is particularly resonant for communities who have been on the receiving end of discriminatory data-driven policies historically, such as the practice of redlining [14, 15]. Other examples include community maps—made to identify problematic properties or an assertion of land rights—being reused by others to identify opportunities for redevelopment or exploitation [16]. Thus, reusing a seemingly public dataset could run counter to the original privacy intents of those who created it and raise questions about whether it represents responsible big data research.

Situate and contextualize your data to anticipate privacy breaches and minimize harm. The availability or perceived publicness of data does not guarantee lack of harm, nor does it mean that data creators consent to researchers using their data.

## 3. Guard against the reidentification of your data

It is problematic to assume that data cannot be reidentified. There are numerous examples of researchers with good intentions and seemingly good methods failing to anonymize data sufficiently to prevent the later identification of specific individuals [17]; in other cases, these efforts were extremely superficial [18, 19]. When datasets thought to be anonymized are combined with other variables, it may result in unexpected reidentification, much like a chemical reaction resulting from the addition of a final ingredient.

While the identificatory power of birthdate, gender, and zip code is well known [20], there are a number of other parameters—particularly the metadata associated with digital activity—that may be as or even more useful for identifying individuals [21]. Surprising to many, unlabeled network graphs—such as location and movement, DNA profiles, call records from mobile phone data, and even high-resolution satellite images of the earth—can be used to reidentify people [22]. More important than specifying the variables that allow for reidentification, however, is the realization that it is difficult to recognize these vulnerable points a priori [23]. Factors discounted today as irrelevant or inherently harmless—such as battery usage—may very well prove to be a significant vector of personal identification tomorrow [24]. For example, the addition of spatial location can turn social media posts into a means of identifying home location [25], and Google's reverse image search can connect previously separate personal activities—such as dating and professional profiles—in unanticipated ways [26]. Even data about groups—"aggregate statistics"—can have serious implications if they reveal that certain communities, for example, suffer from stigmatized diseases or social behavior much more than others [27].

Identify possible vectors of reidentification in your data. Work to minimize them in your published results to the greatest extent possible.

## 4. Practice ethical data sharing

For some projects, sharing data is an expectation of the human participants involved and thus a key part of ethical research. For example, in rare genetic disease research, biological samples are shared in the hope of finding cures, making dissemination a condition of participation. In other projects, questions of the larger public good—an admittedly difficult to define category—provide compelling arguments for sharing data, e.g., the NIH-sponsored database of Genotypes and Phenotypes (dbGaP), which makes deidentified genomic data widely available to researchers, democratizing access, or the justice claim made by the Institute of Medicine about the value of mandating that individual-level data from clinical trials be shared among researchers [28]. Asking participants for broad, as opposed to narrowly structured consent for downstream data management makes it easier to share data. Careful research design and guidance from IRBs can help clarify consent processes. However, we caution that even when broad consent was obtained upfront, researchers should consider the best interests of the human participant, proactively considering the likelihood of privacy breaches and reidentification issues. This is of particular concern for human DNA data, which is uniquely identifiable.

These types of projects, however—in which rules of use and sharing are well governed by informed consent and right of withdrawal—are increasingly the exception rather than the rule for big data. In our digital society, we are followed by data clouds composed of the trace elements of daily life—credit card transactions, medical test results, closed-circuit television (CCTV) images and video, smart phone apps, etc.—collected under mandatory terms of service rather than responsible research design overseen by university compliance officers. While we might wish to have the standards of informed consent and right of withdrawal, these informal big data sources are gathered by agents other than the researcher—private software companies, state agencies, and telecommunications firms. These data are only accessible to researchers after their creation, making it impossible to gain informed consent a priori, and contacting the human participants retroactively for permission is often forbidden by the owner of the data or is impossible to do at scale.

Of course, researchers within software companies and state institutions collecting these data have a special responsibility to address the terms under which data are collected; but that does not exempt the end-user of shared data. In short, the burden of ethical use (see Rules 1 to 3) and sharing is placed on the researcher, since the terms of service under which the human subjects' data were produced can often be extremely broad with little protection for breaches of privacy. In these circumstances, researchers must balance the requirements from funding agencies to share data [29] with their responsibilities to the human beings behind the data they acquired. A researcher needs to inform funding agencies about possible ethical concerns before the research begins and guard against reidentification before sharing.

Share data as specified in research protocols, but proactively address concerns of potential harm from informally collected big data.

## 5. Consider the strengths and limitations of your data; big does not automatically mean better

In order to do both accurate and responsible big data research, it is important to ground datasets in their proper context including conflicts of interests. Context also affects every stage of research: from data acquisition, to cleaning, to interpretation of findings, and dissemination of the results. During the step of data acquisition, it is crucial to understand both the source of the data and the rules and regulations with which they were gathered. This is especially important in cases of research conducted in relatively loose regulatory environments, in which use of answers to research questions may conflict with the expectations of those who provided the data. One possible approach might be the ethical norms employed to track the provenance of artifacts, often in cooperation and collaboration with the communities from which they come (e.g., archaeologists working in indigenous communities to determine the disposition of material culture). In a similar manner, computer scientists use data lineage techniques to track the evolution of a dataset and often to trace bugs in the data.

Being mindful of the data's context provides the foundation for clarifying when your data and analysis are working and when they are not. While it is tempting to interpret findings based on big data as a clear outcome, a key step within scientific research is clearly articulating what data or an indicator represent and what they do not. Are your findings as clear-cut if your interpretation of a social media posting switches from a recording of fact to the performance of a social identity? Given the messy, almost organic nature of many datasets derived from social actions, it is fundamental that researchers be sensitive to the potential multiple meanings of data.

For example, is a Facebook post or an Instagram photo best interpreted as an approval/disapproval of a phenomenon, a simple observation, or an effort to improve status within a friend network? While any of these interpretations are potentially valid, the lack of context makes it even more difficult to justify the choice of one understanding over another. Reflecting on the potential multiple meanings of data fosters greater clarity in research hypotheses and also makes researchers aware of the other potential uses of their data. Again, the act of interpretation is a human process, and because the judgments of those (re)using your data may differ from your own, it is essential to clarify both the strengths and shortcomings of the data.

Document the provenance and evolution of your data. Do not overstate clarity; acknowledge messiness and multiple meanings.

## 6. Debate the tough, ethical choices

Research involving human participants at federally funded institutions is governed by IRBs charged with preventing harm through well-established procedures and are familiar to many researchers. IRBs, however, are not the sole arbiter of ethics; many ethical issues involving big data are outside of their governance mandate. Precisely because big data researchers often encounter situations that are foreign to or outside of the mandate of IRBs, we emphasize the importance of debating the issues within groups of peers.

Rather than a bug, the lack of clear-cut solutions and governance protocols should be more appropriately understood as a feature that researchers should embrace within their own work. Discussion and debate of ethical issues is an essential part of professional development—both within and between disciplines—as it can establish a mature community of responsible practitioners. Bringing these debates into coursework and training can produce peer reviewers who are particularly well placed to raise these ethical questions and spur recognition of the need for these conversations.

A precondition of any formal ethics rules or regulations is the **capacity** to have such open-ended debates. As digital social scientist and ethicist Annette Markham [30] writes, "we can make [data ethics] an easier topic to broach by addressing ethics as being about choices we make at critical junctures; choices that will invariably have impact." Given the nature of big data, bringing technical, scientific, social, and humanistic researchers together on projects enables this debate to emerge as a strength because, if done well, it provides the means to understand the ethical issues from a range of perspectives and disrupt the silos of disciplines

[31]. There are a number of good models for interdisciplinary ethics research, such as the trainings offered by the Science and Justice research center at the University of California, Santa Cruz [32] and Values in Design curricula [33]. Research ethics consultation services, available at some universities as a result of the Clinical and Translational Science Award (CTSA) program of the National Institutes of Health (NIH), can also be resources for researchers [34].

Some of the better-known "big data" ethical cases—i.e., the Facebook emotional contagion study [35]—provide extremely productive venues for cross-disciplinary discussions. Why might one set of scholars see this as a relatively benign approach while other groups see significant ethical shortcomings? Where do researchers differ in drawing the line between responsible and irresponsible research and why? Understanding the different ways people discuss these challenges and processes provides an important check for researchers, especially if they come from disciplines not focused on human subject concerns.

Moreover, the high visibility surrounding these events means that (for better or worse) they represent the "public" view of big data research, and becoming an active member of this conversation ensures that researchers can give voice to their insights rather than simply being at the receiving end of policy decisions. In an effort to help these debates along, the Council for Big Data, Ethics, and Society has produced a number of case studies focused specifically on big data research and a white paper with recommendations to start these important conversations (http://bdes.datasociety.net/output/).

Engage your colleagues and students about ethical practice for big data research.

## 7. Develop a code of conduct for your organization, research community, or industry

The process of debating tough choices inserts ethics directly into the workflow of research, making "faking ethics" as unacceptable as faking data or results. Internalizing these debates, rather than treating them as an afterthought or a problem to outsource, is key for successful research, particularly when using trace data produced by people. This is relevant for all research including those within industry who have privileged access to the data streams of digital daily life. Public attention to the ethical use of these data should not be avoided; after all, these datasets are based on an infrastructure that billions of people are using to live their lives, and there is a compelling public interest that research is done responsibly.

One of the best ways to cement this in daily practice is to develop codes of conduct for use in your organization or research community and for inclusion in formal education and ongoing training. The codes can provide guidance in peer review of publications and in funding consideration. In practice, a highly visible case of unethical research brings problems to an entire field, not just to those directly involved. Moreover, designing codes of conduct makes researchers more successful. Issues that might otherwise be ignored until they blow up—e.g., Are we abiding by the terms of service or users' expectations? Does the general public consider our research "creepy"? [13]—can be addressed thoughtfully rather than in a scramble for damage control. This is particularly relevant to public-facing private businesses interested in avoiding potentially unfavorable attention.

An additional and longer-term advantage of developing codes of conduct is that it is clear that change is coming to big data research. The NSF funded the Council for Big Data, Ethics, and Society as a means of getting in front of a developing issue and pending regulatory changes within federal rules for the protection of human subjects that are currently under review [1]. Actively developing rules for responsible big data research within a research community is a key way researchers can join this ongoing process.

Establish appropriate codes of ethical conduct within your community. Make industry researchers and representatives of affected communities active contributors to this process.

## 8. Design your data and systems for auditability

Although codes of conduct will vary depending on the topic and research community, a particularly important element is designing data and systems for auditability. Responsible internal auditing processes flow easily into audit systems and also keep track of factors that might contribute to problematic outcomes. Developing automated testing processes for assessing problematic outcomes and mechanisms for auditing other's work during review processes can help strengthen research as a whole. The goal of auditability is to clearly document when decisions are made and, if necessary, backtrack to an earlier dataset and address the issue at the root (e.g., if strategies for anonymizing data are compromised).

Designing for auditability also brings direct benefits to researchers by providing a mechanism for double-checking work and forcing oneself to be explicit about decisions, increasing understandability and replicability. For example, many types of social media and other trace data are unstructured, and answers to even basic questions such as network ties, location, and randomness depend on the steps taken to collect and collate data. Systems of auditability clarify how different datasets (and the subsequent analysis) differ from each other, aiding understanding and creating better research.

Plan for and welcome audits of your big data practices.

## 9. Engage with the broader consequences of data and analysis practices

It is also important for responsible big data researchers to think beyond the traditional metrics of success in business and the academy. For example, the energy demands for digital daily life, a key source of big data for social science research, are significant in this era of climate change [36]. How might big data research lessen the environmental impact of data analytics work? For example, should researchers take the lead in asking cloud storage providers and data processing centers to shift to sustainable and renewable energy sources? As important and publicly visible users of the cloud, big data researchers collectively represent an interest group that could rally behind such a call for change.

The pursuit of citations, reputation, or money is a key incentive for pushing research forward, but it can also result in unintended and undesirable outcomes. In contrast, we might ask to what extent is a research project focused on enhancing the public good or the underserved of society? Are questions about equity or promoting other public values being addressed in one's data streams, or is a big data focus rendering them invisible or irrelevant to your analysis [37]? How can increasingly vulnerable yet fundamentally important public resources—such as state-mandated cancer registries—be protected? How might research aid or inhibit different business and political actors? While all big data research need not take up social and cultural questions, a fundamental aim of research goes beyond understanding the world to considering ways to improve it.

Recognize that doing big data research has societal-wide effects.

## 10. Know when to break these rules

The final (and counterintuitive) rule is the charge to recognize when it is appropriate to stray from these rules. For example, in times of natural disaster or a public health emergency, it may be important to temporarily put aside questions of individual privacy in order to serve a larger public good. Likewise, the use of genetic or other biological data collected without informed consent might be vital in managing an emerging disease epidemic.

Moreover, be sure to review the regulatory expectations and legal demands associated with protection of privacy within your dataset. After all, this is an exceedingly slippery slope, so before following this rule (to break others), be cautious that the "emergency" is not simply a convenient justification. The best way to ensure this is to build experience in engaging in the tough debates (Rule 6), constructing codes of conduct (Rule 7), and developing systems for auditing (Rule 8). The more mature the community of researchers is about their processes, checks, and balances, the better equipped it is to assess when breaking the rules is acceptable. It may very well be that you do not come to a final clear set of practices. After all, just as privacy is not binary (Rule 2), neither is responsible research. Ethics is often about finding a good or better, but not perfect, answer, and it is important to ask (and try to answer) the challenging questions. Only through this engagement can a culture of responsible big data research emerge.

Understand that responsible big data research depends on more than meeting checklists.

## Conclusion

The goal of this set of ten rules is to help researchers do better work and ultimately become more successful while avoiding larger complications, including public mistrust. To achieve this, however, scholars must shift from a mindset that is rigorous when focused on techniques and methodology and naïve when it comes to ethics. Statements to the effect that "Data is [sic] already public" [38] are unjustified simplifications of much more complex data ecosystems embedded in even more complex and contingent social practices. Data are people, and to maintain a rigorously naïve definition to the contrary [18] will end up harming research efforts in the long run as pushback comes from the people whose actions and utterances are subject to analysis.

In short, responsible big data research is not about preventing research but making sure that the work is sound, accurate, and maximizes the good while minimizing harm. The problems and choices researchers face are real, complex, and challenging and so too must be our response. We must treat big data research with the respect that it deserves and recognize that unethical research undermines the production of knowledge. Fantastic opportunities to better understand society and our world exist, but with these opportunities also come the responsibility to consider the ethics of our choices in the everyday practices and actions of our research. The Council for Big Data, Ethics, and Society (http://bdes.datasociety.net/) provides an initial set of case studies, papers, and even ten simple rules for guiding this process; it is now incumbent on you to use and improve these in your research.

## Acknowledgments

This article also benefitted from the input of Geoff Bowker and Helen Nissenbaum.

## References

1. Metcalf J, boyd d, Keller E. Perspectives on Big Data, Ethics, and Society. Council for Big Data, Ethics, and Society. 2016. http://bdes.datasociety.net/council-output/perspectives-on-big-data-ethics-and-society/. Accessed 31 May 2016.
2. Wu HY, Rubinstein M, Shih E, Guttag JV, Durand F, Freeman WT. Eulerian video magnification for revealing subtle changes in the world. Eulerian Video Magnification for Revealing Subtle Changes in the World. ACM Transactions on Graphics. 2012; 31(4).

3. Crawford K, Schultz J. Big data and due process: Toward a framework to redress predictive privacy harms. BCL Rev. 2014; 55: 93–128.
4. Barocas S, Selbst AD. Big data's disparate impact. California Law Review. 2016; 104(3): 671–732.
5. Danyllo WA, Alisson VB, Alexandre ND, Moacir LM, Jansepetrus BP, Oliveira RF. Identifying relevant users and groups in the context of credit analysis based on data from Twitter. InCloud and Green Computing (CGC), 2013 Third International Conference on 2013 Sep 30 (pp. 587–592). IEEE.
6. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. Pro Publica. 23 May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed 4 September 2016.
7. Ingold D, Spencer S. Amazon Doesn't Consider the Race of Its Customers. Should It? Bloomberg.com 21 April 2016. http://www.bloomberg.com/graphics/2016-amazon-same-day/. Accessed 12 June 2016.
8. Hauge MV, Stevenson MD, Rossmo DK, Le Comber SC. Tagging Banksy: using geographic profiling to investigate a modern art mystery. Journal of Spatial Science. 2016; 61(1):185–90.
9. Metcalf J, Crawford K. Where are Human Subjects in Big Data Research? The Emerging Ethics Divide. The Emerging Ethics Divide. Big Data and Society, 2016.
10. Zwitter A. Big data ethics. Big Data & Society. 2014; 1(2).
11. Nissenbaum H. Privacy in context: Technology, policy, and the integrity of social life. Stanford University Press; 2009.
12. Marwick AE. boyd d. Networked privacy: How teenagers negotiate context in social media. New Media & Society. 2014:1461444814543995.
13. Tene O, Polonetsky J. Theory of Creepy: Technology, Privacy and Shifting Social Norms, A. Yale JL & Tech. 2013; 16:59.
14. Massey DS, Denton NA. American apartheid: Segregation and the making of the underclass. Harvard University Press; 1993.
15. Davidow B. Redlining for the 21st Century. The Atlantic. 5 March 2014. http://www.theatlantic.com/business/archive/2014/03/redlining-for-the-21st-century/284235/. Accessed 31 May 2016.
16. Young JC, Gilmore MP. Subaltern empowerment in the Geoweb: Tensions between publicity and privacy. Antipode. 2014; 46(2):574–91.
17. Barbaro M, Zeller T, Hansell S. A face is exposed for AOL searcher no. 4417749. New York Times. 2006 Aug 9;9.
18. Cox J. 70,000 OkCupid Users Just Had Their Data Published. Motherboard. 12 May 2016. http://motherboard.vice.com/read/70000-okcupid-users-just-had-their-data-published. Accessed 12 June 2016.
19. Pandurangan V. On Taxis and Rainbows: Lessons from NYC's improperly anonymized taxi logs. Medium. 2014. https://medium.com/@vijayp/of-taxis-and-rainbows-f6bc289679a1. Accessed 10 November 2015.
20. Sweeney L. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002; 10(05):557–70.
21. Zimmer M. "But the data is already public": on the ethics of research in Facebook. Ethics and information technology. 2010; 12(4):313–25.
22. Kloumann IM, Kleinberg JM. Community membership identification from small seed sets. InProceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014 Aug 24 (pp. 1366–1375). ACM.
23. Narayanan A, Huey J, Felten EW. A precautionary approach to big data privacy. In Data protection on the move 2016 (pp. 357–385). Springer Netherlands.
24. Michalevsky Y, Schulman A, Veerapandian GA, Boneh D, Nakibly G. Powerspy: Location tracking using mobile device power analysis. In24th USENIX Security Symposium (USENIX Security 15) 2015 (pp. 785–800).
25. Shelton T, Poorthuis A, Zook M. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. Landscape and Urban Planning. 2015; 142:198–211.
26. Acquisti A, Gross R, Stutzman F. Face recognition and privacy in the age of augmented reality. Journal of Privacy and Confidentiality. 2014; 6(2): 1–20.
27. Homer N, Merriman B, Nelson SF. BFAST: an alignment tool for large scale genome resequencing. PLoS ONE. 2009; 4(11):e7767. https://doi.org/10.1371/journal.pone.0007767 PMID: 19907642
28. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. Jama. 2015; 313(8):793–4. https://doi.org/10.1001/jama.2015.292 PMID: 25594500

29. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten simple rules for the care and feeding of scientific data. PLoS Comput Biol.; 10(4).

30. Markham A. OKCupid data release fiasco: It's time to rethink ethics education. Medium.Points 18 May 2016. https://points.datasociety.net/okcupid-data-release-fiasco-ba0388348cd#.g4ofbpnc6. Accessed 12 June 2016.

31. Ford H. Big Data and Small: Collaborations between ethnographers and data scientists. Big Data & Society. 2014; 1(2):

32. Science & Justice Research Center (Collaborations Group. Experiments in collaboration: interdisciplinary graduate education in science and justice. PLoS Biol. 2013 Jul 30; 11(7):e1001619.

33. Knobel C, Bowker GC. Values in design. Communications of the ACM. 2011; 54(7):26–8.

34. Cho MK, Tobin SL, Greely HT, McCormick J, Boyce A, Magnus D. Research ethics consultation: The Stanford experience. IRB. 2008;(6):1–6. PMID: 19119757

35. boyd d. Untangling research and practice: What Facebook's "emotional contagion" study teaches us. Research Ethics. 2016; 12(1):4–13.

36. Cook G, Dowdall T, Pomerantz D, Wang Y. Clicking clean: how companies are creating the green internet. Greenpeace Inc., Washington, DC. 2014. http://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/planet3/PDFs/clickingclean.pdf

37. Zook MA, Graham M. Mapping DigiPlace: geocoded Internet data and the representation of place. Environment and Planning B: Planning and Design. 2007 Jun 1; 34(3):466–82.

38. Zimmer M. OkCupid Study Reveals the Perils of Big-Data Science. Wired. 14 May 2016. https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/. Accessed 12 June 2016.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Digital Data Storage

**Edmund M. Hart[1]*, Pauline Barmby[2], David LeBauer[3], François Michonneau[4,5], Sarah Mount[6], Patrick Mulrooney[7], Timothée Poisot[8], Kara H. Woo[9], Naupaka B. Zimmerman[10], Jeffrey W. Hollister[11]**

**1** University of Vermont, Department of Biology, Burlington, Vermont, United States of America, **2** University of Western Ontario, Department of Physics and Astronomy, London, Canada, **3** University of Illinois at Urbana-Champaign, National Center for Supercomputing Applications and Institute for Genomic Biology, Urbana, Illinois, United States of America, **4** University of Florida, iDigBio, Florida Museum of Natural History, Gainesville, Florida, United States of America, **5** University of Florida, Whitney Laboratory for Marine Bioscience, Gainesville, Florida, United States of America, **6** King's College London, Department of Informatics, London, United Kingdom, **7** University of California at San Diego, San Diego Supercomputer Center, San Diego, California, United States of America, **8** Université de Montréal, Département de Sciences Biologiques, Montreal, Canada, **9** Washington State University, Center for Environmental Research, Education, and Outreach, Pullman, Washington, United States of America, **10** University of Arizona, School of Plant Sciences, Tucson, Arizona, United States of America, **11** US Environmental Protection Agency, Atlantic Ecology Division, Narragansett, Rhode Island, United States of America

* edmund.m.hart@gmail.com



**Citation:** Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. (2016) Ten Simple Rules for Digital Data Storage. PLoS Comput Biol 12(10): e1005097. doi:10.1371/journal.pcbi.1005097

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** October 20, 2016



**Funding:** DL was funded by the Energy Biosciences institute and the National Center for Supercomputing Applications. FM was funded by iDigBio (Integrated Digitized Biocollections), and therefore this material is based upon work supported by the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). KHW was funded by Washington State University. NBZ was funded by the Gordon and Betty Moore Foundation through Grant GBMF 2550.03 to the Life Sciences Research Foundation. PB was funded by the Natural Sciences & Engineering Research Council of Canada and the Academic Development Fund of the University of Western Ontario. TP was funded by an NSERC Discovery Grant, and a Start-

## Introduction

Data is the central currency of science, but the nature of scientific data has changed dramatically with the rapid pace of technology. This change has led to the development of a wide variety of data formats, dataset sizes, data complexity, data use cases, and data sharing practices. Improvements in high-throughput DNA sequencing, sustained institutional support for large sensor networks [1,2], and sky surveys with large-format digital cameras [3] have created massive quantities of data. At the same time, the combination of increasingly diverse research teams [4] and data aggregation in portals (e.g., for biodiversity data, GBIF.org or iDigBio) necessitates increased coordination among data collectors and institutions [5,6]. As a consequence, "data" can now mean anything from petabytes of information stored in professionally maintained databases, to spreadsheets on a single computer, to handwritten tables in lab notebooks on shelves. All remain important, but data curation practices must continue to keep pace with the changes brought about by new forms of data and new data collection and storage practices.

While much has been written about both the virtues of data sharing [7,8] and the best practices to do so [9,10], data storage has received comparatively less attention. Proper storage is a prerequisite to sharing, and indeed inadequate storage contributes to the phenomenon of data decay or to "data entropy," in which data, whether publicly shared or not, becomes less accessible through time [11–14]. Best practices for data storage often begin and end with this statement: "Deposit your data in a community standard repository." This is good advice, especially considering your data is most likely to be reused if it is available on a community site. Community repositories can also provide guidance for best practices. As an example, if you are archiving sequencing data, a repository such as those run by the National Center for Biotechnology Information (NCBI) (e.g., GenBank) not only provides a location for data archival but also encourages a set of practices related to consistent data formatting and the inclusion of appropriate metadata. However, data storage policies are highly variable between repositories [15]. A data management plan utilizing best practices across all stages of the data life cycle will facilitate transition from local storage to repository [16]. Similarly, having such a plan can facilitate

Up grant from the Université de Montréal. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

transition from repository to repository if funding runs out or requirements change. Good storage practices are important even (or especially) in cases when data may not fit with an existing repository, when only derived data products (versus raw data) are suitable for archiving, or when an existing repository may have lax standards.

This article describes ten simple rules for digital data storage that grew out of a long discussion among instructors for the Software and Data Carpentry initiatives [17,18]. Software and Data Carpentry instructors are scientists from diverse backgrounds who have encountered a variety of data storage challenges and are active in teaching other scientists best practices for scientific computing and data management. Thus, this paper represents a distillation of collective experience, and hopefully will be useful to scientists facing a variety of data storage challenges. We additionally provide a glossary of common vocabulary for readers who may not be familiar with particular terms.

## Rule 1: Anticipate How Your Data Will Be Used

One can avoid most of the troubles encountered during the analysis, management, and release of data by having a clear roadmap of what to expect before data acquisition starts. For instance:

- How will the raw data be received? Are they delivered by a machine or software, or typed in?
- What is the format expected by the software used for analysis?
- Is there a community standard format for this type of data?
- How much data will be collected, and over what period of time?

The answers to these questions can range from simple cases (e.g., sequencing data stored in the FASTA format, which can be used "as is" throughout the analysis), to experimental designs involving multiple instruments, each with its own output format and processing conventions. Knowing the state in which the data needs to be at each step of the analysis can help to (i) identify software tools to use in converting between data formats, (ii) orient technological choices about how and where the data should be stored, and (iii) rationalize the analysis pipeline, making it more amenable to re-use [19].

Also key is the ability to estimate the storage volume needed to store the data, both during and after the analysis. The required strategy will differ for datasets of varying size. Smaller datasets (e.g., a few megabytes in size) can be managed locally with a simple data management plan, whereas larger datasets (e.g., gigabytes to petabytes) will in almost all cases require careful planning and preparation (Rule 10).

Lastly, early consideration and planning should be given to the metadata of the project. A plan should be developed early as to what metadata will be collected and how it will be maintained and stored (Rule 7). Also be sure to consider community software tools that can facilitate metadata curation and repository submission. Examples in the biological sciences include *Morpho* for ecological metadata [20] and *mothur* [21] for submitting to NCBI's Sequence Read Archive.

## Rule 2: Know Your Use Case

Well-identified use cases make data storage easier. Ideally, prior to beginning data collection, researchers should be able to answer the following questions:

- Should the raw data be archived (Rule 3)?

- Should the data used for analysis be prepared once or re-generated from the raw data each time (and what difference would this choice make for storage, computing requirements, and reproducibility)?
- Can manual corrections be avoided in favor of programmatic or self-documenting approaches (e.g., Jupyter notebook or R markdown)?
- How will changes to the data be tracked, and where will these tracked changes be logged?
- Will the final data be released, and if so, in what format?
- Are there restrictions or privacy concerns associated with the data (e.g., survey results with personally identifiable information [PII], threatened species, or confidential business information)?
- Will institutional validation be required prior to releasing the data?
- Does the funding agency mandate data deposition in a publicly available archive, and if so, when, where, and under what license?
- Does the target journal mandate data deposition?

None of these questions have universal answers, nor are they the only questions to ask before starting data acquisition. But knowing the what, when, and how of your use of the data will bring you close to a reliable roadmap on how to handle data from acquisition through publication and archival.

## Rule 3: Keep Raw Data Raw

Since analytical and data processing procedures improve or otherwise change over time, having access to the "raw" (unprocessed) data can facilitate future re-analysis and analytical reproducibility. As processing algorithms improve and computational power increases, new analyses will be enabled that were not possible at the time of the original work. If only derived data are stored, it can be difficult for other researchers to confirm analytical results, to assess the validity of statistical models, or to directly compare findings across studies.

Therefore, data should always be kept in raw format whenever possible (within the constraints of technical limitations). In addition to being the most appropriate way to ensure transparency in analysis, having the data stored and archived in their original state gives a common point of reference for derivative analyses. What constitutes sufficiently "raw" data is not always clear (e.g., ohms from a temperature sensor or images of an Illumina sequencing flowcell are generally not archived after the initial processing). Yet the spirit of this rule is that data should be as "pure" as possible when they are stored. If derivations occur, they should be documented by also archiving relevant code and intermediate datasets.

A cryptographic hash (e.g., SHA or MD5) of the raw data should be generated and distributed with the data. These hashes ensure that the dataset has not suffered any silent corruption and/or manipulation while being stored or transferred (see Internet2 Silent Data Corruption). For large enough datasets, the likelihood of silent data corruption is high. This technique has been widely used by many Linux distributions to distribute images and has been very effective with minimal effort.

## Rule 4: Store Data in Open Formats

To maximize accessibility and long-term value, it is preferable to store data in formats that have freely available specifications. The appropriate file type will depend on the data being stored (e.g., numeric measurements, text, images, video), but the key idea is that accessing data should not require proprietary software, hardware, or purchase of a commercial license. Proprietary formats change, maintaining organizations go out of business, and changes in license fees make access to data in proprietary formats unaffordable and risky for end-users. Examples of open data formats include comma-separated values (CSV) for tabular data, hierarchical data format (HDF) [22] and NetCDF [23] for hierarchically structured scientific data, portable network graphics (PNG) for images, KML (or other Open Geospatial Consortium [OGC] format) for spatial data, and extensible markup language (XML) for documents. Examples of closed formats include DWG for AutoCAD drawings, Photoshop document (PSD) for bitmap images, Windows Media Audio (WMA) for audio recording files, and Microsoft Excel (XLS) for tabular data. Even if day-to-day processing uses closed formats (e.g., due to software requirements), data being stored for archival purposes should be stored in open formats. This is generally not prohibitive; most closed-source software products enable users to export data to an open format.

Not only should data be stored in an open format but it should also be stored in a format that computers can easily use for processing. This is especially crucial as datasets become larger. Making data easily usable is best achieved by using standard data formats that have open specifications (e.g., CSV, XML, JSON, HDF5), or by using databases. Such data formats can be handled by a variety of programming languages, as efficient and well-tested libraries for parsing them are typically available. These standard data formats also ensure interoperability, facilitate re-use, and reduce the chances of data loss or mistakes being introduced during conversion between formats. Examples of machine-readable open formats that would not be easy to process include data included in the text of a PDF file or scanned images of tabular data from a paper source.

## Rule 5: Data Should Be Structured for Analysis

To take full advantage of data, it can be useful for it to be structured in a way that makes use, interpretation, and analysis easy. One such structure for data stores each variable as a column, each observation as a row, and each type of observational unit as a table (Fig 1). The technical term for this structure is "Codd's 3rd normal form," but it has been made more accessible as the concept of tidy data [24]. When data is organized in this way, the duplication of information is reduced and it is easier to subset or summarize the dataset to include the variables or observations of interest.

One axiom about the structure of data and code holds that one should "write code for humans, write data for computers" [25]. When data can be easily imported and manipulated using familiar software (whether via a scripting language, a spreadsheet, or any other computer program that can import these common files), data becomes easier to re-use. Furthermore, having the source code for the software doing the analysis available provides provenance for how the data is processed and analyzed. This makes analysis more transparent, since all assumptions about the structure of the data are implicitly stated in the source code. This also enables extraction of the analyses performed, their reproduction, and their modification.

Interoperability is facilitated when variable names are mapped to existing data standards. For instance, for biodiversity data, the Darwin Core Standard provides a set of terms that describe observations, specimens, samples, and related information for a taxa. For earth science and ecosystem models and data, the Climate Forecasting Conventions are widely adopted, such that a large ecosystem of software and data products exist to reduce the technical burden of reformatting and reusing large and complex data. Because each term in such standards is clearly defined and documented, each dataset can use the terms consistently; this facilitates

A

Untidy Data

| species | habitat | weight | length | latitude/longitude | date |
|---|---|---|---|---|---|
| Alligator mississippiensis | swamp | 431 lb | 4 ft 2 | 29.531,-82.184 | Sept 15, 2015 |
| Puma concolor | forest | 125 lb | 2.2m | 29.125,-81.682 | 08/10/2015 |
| Ursus americanus | forest | 88 kg | 133 cm | N29°7'30"/W81°40'55.2" | 07-13-2015 |

B

Tidy Data

meta-data | data

| species_code | date | station_code | weight_kg | length_cm |
|---|---|---|---|---|
| TSN 551771 | 2015-09-15 | 1 | 196 | 127 |
| TSN 55247 | 2015-08-10 | 2 | 57 | 220 |
| TSN 180544 | 2015-07-13 | 2 | 88 | 133 |

| station_code | habitat | latitude | longitude |
|---|---|---|---|
| 1 | swamp | 29.531 | -82.184 |
| 2 | forest | 29.125 | -81.682 |

| species_code | class | genus | species |
|---|---|---|---|
| TSN 551771 | Reptilia | Alligator | mississippiensis |
| TSN 55247 | Mammalia | Puma | concolor |
| TSN 180544 | Mammalia | Ursus | americanus |

**Fig 1. Example of an untidy dataset (A) and its tidy equivalent (B).** Dataset A is untidy because it mixes observational units (species, location of observations, measurements about individuals), the units are mixed and listed with the observations, more than one variable is listed (both latitude and longitude for the coordinates, and genus and species for the species names), and several formats are used in the same column for dates and geographic coordinates. Dataset B is an example of a tidy version of dataset A that reduces the amount of information that is duplicated in each row, limiting chances of introducing mistakes in the data. By having species in a separate table, they can be identified uniquely using the Taxonomic Serial Number (TSN) from the Integrated Taxonomic Information System (ITIS), and it makes it easy to add information about the classification of these species. It also allows researchers to edit the taxonomic information independently from the table that holds the measurements about the individuals. Unique values for each observational unit facilitate the programmatic combination of information using "join" operations. With this example, if the focus of the study for which these data were collected is based upon the size measurements of the individuals (weight and length), information about "where," "when," and "what" animals were measured can be considered metadata. Using the tidy format makes this distinction clearer.

doi:10.1371/journal.pcbi.1005097.g001

data sharing across institutions, applications, and disciplines. With machine-readable, standards-compliant data, it becomes easier to build an Application Programming Interface (API) to query the dataset and retrieve a subset of interest, as outlined in Rule 10.

## Rule 6: Data Should Be Uniquely Identifiable

To aid reproducibility, the data used in a scientific publication should be uniquely identifiable. Ideally, datasets should have a unique identifier such as a Digital Object Identifier (DOI), Archival Resource Key (ARK), or a persistent URL (PURL). An increasing number of online services, such as Figshare, Zenodo, or DataOne, are able to provide these. Institutional initiatives also exist and are known to your local librarians. Some repositories may require specific identifiers, and these could change with time. For instance, NCBI sequence data will in the future only be identified by "accession.version" IDs. The "GI" identifiers (in use since 1994) will be retired in late 2016 [26].

Even as identifier standards may change over time, datasets can evolve over time as well. In order to distinguish between different versions of the same data, each dataset should have a distinct name, which includes a version identifier. A simple way to do this is to use date stamps as part of the dataset name. Using the ISO 8601 standard avoids regional ambiguities: it mandates the date format `YYYY-MM-DD` (i.e. from largest time unit to smallest). For example, the date "February 1, 2015," while written as `01-02-2015` in the UK and `02-01-2015` in the US, is unambiguous (`2015-02-01`) under this standard.

Semantic versioning is a richer approach to solving the same problem [27]. The CellPack datasets are an example of this [28]. A semantic version number takes the form: `Major.Minor.Patch`, e.g., `0.2.7`. The *major version* numbers should be incremented (or bumped) when a dataset scheme has been updated or some other change is made that is not compatible with previous versions of the data with the same major version number. This means that an experiment using version `1.0.0` of the dataset may not run on version `2.0.0` without changes to the data analysis. The *minor version* should be bumped when a change has been made that is compatible with older versions of the data with the same major version. This means that any analysis that can be performed on version `1.0.0` of the data is repeatable with version `1.1.0` of the data. For example, adding a new year in a temporal survey will result in a bump in the minor version. The *patch version* number is bumped when typos or bugs have been fixed. For example version `1.0.1` of a dataset may fix a typo in version `1.0.0`.

## Rule 7: Link Relevant Metadata

Metadata is the contextual information required to interpret data (Fig 1) and should be clearly defined and tightly integrated with data. The importance of metadata for context, reusability, and discovery has been written about at length in guides for data management best practices [9,13,29].

Metadata should be as comprehensive as possible, using standards and conventions of a discipline, and should be machine-readable. Metadata should always accompany a dataset, wherever it is stored, but the best way to do this depends on the format of the data. Text files can contain metadata in well-defined text files such as XML or JSON. Some file formats are self-documenting; for example, NetCDF, HDF5, and many image file formats allow for embedded metadata [22,23]. In a relational database, metadata tables can be clearly labeled and linked to the data. Ideally, a schema will be provided that also shows the linkages between data tables and metadata tables. Another—simpler—scenario is a set of flat (non-hierarchical) text files—in this case a semantically versioned, compressed archive should be created that includes metadata.

Whatever format is used for archiving, the goal should be to make the link between metadata and data as clear as possible. The best approach is dependent on the archiving plan used, but even if the dataset is archived solely for personal use, metadata will provide crucial context for future reuse.

## Rule 8: Adopt the Proper Privacy Protocols

In datasets for which privacy is important, be sure to have a plan in place to protect data confidentiality. You should consider the different data stakeholders when developing privacy protocols for your data storage. These stakeholders include funding agencies, human subjects or entities, collaborators, and yourself. Both the United States National Science Foundation and

National Institutes of Health have data sharing policies in their grant guidelines to prevent sharing personally identifiable information and to anonymize data on human subjects.

In small datasets, a lookup table (protecting PII by removing it and replacing it with a unique ID that maps to the sensitive data in an external dataset) is enough to anonymize a minimal amount of personal information. Hashing techniques are susceptible to a number of attacks, and all hashed data can eventually be determined. Famously, New York City officials shared what they thought was anonymized data on cab drivers and over 173 million cab rides. However, it was quickly recognized that the city anonymized the data with a simple MD5 hashing scheme and all 20 GB of data were de-anonymized in a matter of hours [30]. This type of error can be prevented by asking a trusted colleague or security personnel to try to "crack" anonymised data before releasing it publicly. Often the person who has produced the data is not well placed to check the fine details of their own security procedures. If possible, the best solution is to remove any sensitive data that is not required from the dataset prior to distribution.

In more problematic cases, the data itself allows identifiability: this is the case with human genomic data that map directly onto a subject's identity [31]. Methods for dealing with these complex issues at the intersection of data storage and privacy are rapidly evolving and include storing changes against a reference genome to help with privacy and reduce overall data volumes [32,33] and/or bringing computation to data storage facilities instead of vice versa [34]. Having a plan for privacy before data is acquired is important because it can determine or limit how data will be stored.

## Rule 9: Have a Systematic Backup Scheme

Every storage medium can fail, and every failure can result in loss of data. Researchers should therefore back up data at all stages of the research process. Data stored on local computers or institutional servers during the collection and analysis phases of a project should be backed up to other locations to protect against data loss. No backup system is failsafe (see the stories of the Dedoose crash and the near deletion of Toy Story 2), so more than one backup system should be used. Ideally you should have two on-site copies (such as on a computer, an external hard drive, or a tape) and one off-site copy (e.g., cloud storage) [35], with care taken to ensure that the off-site copy is as secure as the on-site copies. Keeping backups in multiple locations additionally protects against data loss due to theft or natural disasters.

Researchers should test their backups regularly to ensure that they are functioning properly. Common reasons for backup failure include:

- faulty backup software
- incorrect configuration (e.g., not backing up sub-directories)
- encryption (e.g., someone encrypted the backups but later lost the password to decrypt them)
- media errors

Consider the backup plan of your selected data repository before publishing your data and if possible, find out about the long-term storage plans of the repository. Many repositories mirror the data they host on multiple machines. Are there plans in place to keep data available if the organization that manages the repository dissolves?

## Rule 10: The Location and Method of Data Storage Depend on How Much Data You Have

The storage method you should choose depends on the size and nature of your data, the cost of storage and access over time, the time it takes to transfer the data, how the data will be used, and any privacy concerns. Data is increasingly generated in the range of many terabytes (TB) by environmental sensors, satellites, automated analytical tools, simulation models, and nucleic acid sequencers. Even larger data-generating machines, like the Large Hadron Collider (LHC) and the Large Scale Synoptic Survey Telescope (LSST), generate many TB per day, rapidly accumulating to petabyte (PB) scale over the course of any particular study. While the cost of storage continues to decrease, the volume of data to be stored impacts the choice of storage methods and locations: for large datasets it is necessary to balance the cost of storage with the time of access and costs of re-generating the data. With new commercial cloud offerings (e.g., Amazon S3) the cost of retrieving the data might exceed the cost of analysis or of re-generating the data from scratch.

When data takes too long to transfer or is costly to store, it can become more efficient to use a computer system for analysis that can directly access and use the data in place instead of first transferring it to a local machine. Inactive data can be put in longer-term storage; this is less expensive, but can take longer to retrieve. Some storage systems automatically migrate "stale" files to longer-term storage. Alternatively, some computing can be done "in the database" or "on disk" via database query languages (e.g., SQL, MapReduce) that perform basic arithmetic, or via the use of procedural languages (e.g., R, Python, C) embedded in the database server. Modern database technologies such as HDFS and Spark allow these computations to be done on data of almost any size. When data is larger than locally available RAM, it can be handled by conducting analyses on a "big memory" node, which most high-performance computing centers have deployed. Relying on tight software/hardware integration, these can allow for the analysis of datasets around 1–4 TB in size. This allows the user to read in and use a large dataset without special tools.

If you regularly only need access to a small subset of your data or need to share it with many collaborators, a web-based API (Application Programming Interface) might be a good solution. Using this method, many users can send requests to an online service that can subset the data, perform in-database computation, and return smaller volumes of data as specific slices. Tools based on online services make it easier to find and download data, and they facilitate analysis via reproducible scripts. However, they can also lead to excessive and careless abuse of resources without proper safeguards in place. The time required to re-download and re-compute results can be reduced by "caching." Caching stores copies of downloads and generated files that are recognized when the same script is run multiple times.

## Further Reading and Resources

Digital data storage is a vast topic; the references given here and elsewhere in this paper provide some starting points for interested readers. For beginning users of scientific data, Data Carpentry offers workshops and resources on data management and analysis, as do the DataONE education modules [36]. For librarians and others who are responsible for data archiving, Data Curation Profiles [37] may be of interest.

## Glossary

### Projects and initiatives

- **Global Biodiversity Information Facility** (GBIF, http://www.gbif.org) provides an international open data infrastructure to publish and disseminate biodiversity information.
- **Integrated Digitized Biocollections** (iDigBio, https://www.idigbio.org) is a project funded by the National Science Foundation that facilitates the digitization of natural history collections and provides data and images for biological specimens.

- **Integrated Taxonomic Information System** (ITIS, http://www.itis.gov) is an international partnership of governmental organizations that aims at providing authoritative taxonomic information for plants, animals, fungi, and microbes.

## File formats

- **Comma-Separated Values** (CSV) and **Tab-Separated Values** (TSV) are plain text file formats used to store tabular data, in which each row is represented by a line in the file and each field (column) is separated by a comma (for CSV) or by the tab character (for TSV).
- **FASTA** is a simple and widely used file format used to represent sequences of nucleotides or amino acids in plain text, making it easy to manipulate these programmatically.
- **Hierarchical Data Format** (HDF) is an open-source binary file format designed to store large amounts of data (and their associated metadata) by providing a hierarchical structure that could be compared to how a hard drive is organized with directories and files. It is maintained by the non-profit HDF Group, a spin-off of the National Center for Supercomputing Applications (NCSA).
- **JavaScript Object Notation** (JSON) is a plain text file format typically used to store arbitrarily structured data in the form of keys and values. It can be used to store non-relational databases, as it does not rely on a tabular data format. In many respects, it has been replacing XML.
- **Network Common Data Form** (NetCDF) is an open-source binary file format designed to store large datasets in array-oriented scientific data, typically used in the geosciences. It is maintained by Unidata, a non-profit member of the University Corporation for Atmospheric Research (UCAR), which is funded by the National Science Foundation.
- **Extensible Markup Language** (XML) is a markup language and the file format used to store documents written with it. It is used to represent arbitrary data structures and is both human and machine-readable.

## Programming and algorithms

- **Web Application Programming Interface** (API) provide ways to programmatically query databases through the internet. They notably allow users to retrieve and work with a small slice of a large dataset.
- **Hadoop Distributed File System** (HDFS) is a Java-based file system in which data is stored in small chunks across multiple redundant nodes.
- **MapReduce** is a style of programming designed to work with large datasets in parallel computing environments. Such programs are composed of a **map** procedure in which the dataset is sliced into several pieces, and a **reduce** procedure in which summary operations are then applied to each of the slices.
- **Secure Hash Algorithm 2** (SHA-2) is a family of Secure Hashing Algorithms used in cryptographic analysis, often to verify the integrity of a file. A cryptographic hash function converts a "message" (e.g., passwords, file content) into an encrypted value. Cryptographic hash functions are easy to compute from the message, but it should be impossible to recover the

message from the output, and any modifications to the message should also modify the output. The SHA algorithms are often used in preference to similar tools such as MD5 (mentioned in Rule 3 and in Rule 8), which are no longer secure. All hashing algorithms are vulnerable to brute force attacks. Key Derivation Function (KDF) implementations like BCrypt and PBKDF2 are considered significantly more secure, but by design more costly to compute.

- Apache **Spark** is an open-source computing platform for querying large datasets in memory, in contrast to on-disk–based methods like MapReduce.
- **Structured Query Language** (SQL) is a programming language used to interact with relational database management systems.

### Hardware

- **mega-, giga-, tera-, petabytes** are units of digital information and are used to measure the size of datasets or the storage media. Originally a byte was the minimum amount of memory needed to store a single character of text in a computer. The prefixes mega-, giga-, tera-, and peta- refer to the international system of units for the multiple of the unit and correspond to $10^6$, $10^9$, $10^{12}$, and $10^{15}$, abbreviated M, G, T, and P, respectively.

### Persistent identifiers

- **Archival Resource Key** (ARK) identifiers are URLs designed to support long-term access to information online.
- **Digital Object Identifier** (DOI) provides unique and persistent identifiers for electronic documents (in particular, journal articles and datasets) on the internet. The uniqueness of the identifiers is guaranteed by a central registry. By dissociating the identifier and the location of the document (i.e., the URL), the DOI can remain fixed even if the location of the digital object it is pointing to changes.
- **Persistent Uniform Resource Locator** (PURL) is a URL used to redirect to the location of an electronic object on the internet. DOI and ARK are examples of implementations of PURL.
- **Uniform Resource Locator** (URL) gives the location of an object on the World Wide Web; the most familiar type of URL is a website address.

## Acknowledgments

We would like to thank G. Wilson and the Software Carpentry instructor community for providing the initial inspiration for this paper and the discussion held at https://github.com/swcarpentry/site/issues/797. We thank K. Hinsen, G. Pugh, and W. Turner for comments on the draft manuscript. Contributions from J. W. Hollister, an employee of the US Environmental Protection Agency (EPA) have not been subjected to the Agency's peer and administrative review. Therefore, the conclusions and opinions drawn are solely those of the author(s) and are not necessarily the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. S. Mount would like to thank Dr. Laurence Tratt for reviewing an early draft of this paper.

## References


1. Reid JG, Carroll A, Veeraraghavan N, Dahdouli M, Sundquist A, English A, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. BMC bioinformatics. 2014; 15: 30. doi: 10.1186/1471-2105-15-30 PMID: 24475911
2. Hampton SE, Strasser C, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, et al. Big data and the future of ecology. Frontiers in Ecology and the Environment. 2013; 130312142848005. doi: 10.1890/120103
3. Eisenstein DJ, others. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. The Astronomical Journal. 2011; 142: 72. doi: 10.1088/0004-6256/142/3/72
4. Adams J. Collaborations: The rise of research networks. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. 2012; 490: 335–6.
5. Fraser LH, Henry HA, Carlyle CN, White SR, Beierkuhnlein C, Cahill JF, et al. Coordinated distributed experiments: an emerging tool for testing global hypotheses in ecology and environmental science. Frontiers in Ecology and the Environment. Ecological Society of America; 2013; 11: 147–155. doi: 10.1890/110279
6. Robertson MAG Tim AND Döring. The GBIF integrated publishing toolkit: Facilitating the efficient publishing of biodiversity data on the internet. PLoS ONE. Public Library of Science; 2014; 9: e102623. doi: 10.1371/journal.pone.0102623 PMID: 25099149
7. Wolkovich EM, Regetz J, O'Connor MI. Advances in global change research require open science by individual researchers. Global Change Biology. 2012; 18: 2102–2110. doi: 10.1111/j.1365-2486.2012.02693.x
8. Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, Cain KE, et al. Troubleshooting public data archiving: suggestions to increase participation. PLoS biology. 2014; 12: e1001779. doi: 10.1371/journal.pbio.1001779 PMID: 24492920
9. White E, Baldridge E, Brym Z, Locey K, McGlinn D, Supp S. Nine simple ways to make it easier to (re) use your data. Ideas in Ecology and Evolution. 2013; 6: 1–10. doi: 10.4033/iee.2013.6b.6.f
10. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten simple rules for the care and feeding of scientific data. PLoS computational biology. 2014; 10: e1003542. doi: 10.1371/journal.pcbi.1003542 PMID: 24763340
11. Pepe A, Goodman A, Muench A, Crosas M, Erdmann C. How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. Golden AA-J, editor. PLoS ONE. 2014; 9: e104798. doi: 10.1371/journal.pone.0104798 PMID: 25165807
12. Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, et al. The availability of research data declines rapidly with article age. Current biology: CB. Elsevier; 2014; 24: 94–7. doi: 10.1016/j.cub.2013.11.014 PMID: 24361065
13. Michener WK, Jones MB. Ecoinformatics: supporting ecology as a data-intensive science. Trends in ecology & evolution. 2012; 27: 85–93. doi: 10.1016/j.tree.2011.11.016 PMID: 22240191
14. Michener WK, Brunt JW, Helly JJ, Kirchner TB, Stafford SG. Nongeospatial metadata for the ecological sciences. Ecological Applications. Eco Soc America; 1997; 7: 330–342. doi: 10.1890/1051-0761(1997)007%5B0330:NMFTES%5D2.0.CO;2
15. Marcial LH, Hemminger BM. Scientific data repositories on the Web: An initial survey. Journal of the American Society for Information Science and Technology. 2010; 61: 2029–2048. doi: 10.1002/asi.21339
16. Michener WK. Ten simple rules for creating a good data management plan. PLoS computational biology. Public Library of Science; 2015; 11: e1004525. doi: 10.1371/journal.pcbi.1004525 PMID: 26492633
17. Wilson G. Software Carpentry: lessons learned. F1000Research. 2014; 3: 62. doi: 10.12688/f1000research.3-62.v1 PMID: 24715981
18. Teal TK, Cranston Ka, Lapp H, White E, Wilson G, Ram K, et al. Data Carpentry: Workshops to Increase Data Literacy for Researchers. International Journal of Digital Curation. 2015; 10: 135–143. doi: 10.2218/ijdc.v10i1.351
19. Roche DG, Kruuk LEB, Lanfear R, Binning SA. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLoS Biology. 2015; 13: e1002295. doi: 10.1371/journal.pbio.1002295 PMID: 26556502
20. Higgins D, Berkley C, Jones M. Managing heterogeneous ecological data using morpho [Internet]. IEEE Comput. Soc; 2002. pp. 69–76. doi: 10.1109/SSDM.2002.1029707

21. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology. Am Soc Microbiol; 2009; 75: 7537–7541. doi: 10.1128/AEM.01541-09 PMID: 19801464
22. Koziol Q, Matzke R. Hdf5: A new generation of hdf: Reference manual and user guide. National Center for Supercomputing Applications, Champaign, Illinois, USA, http://hdf.ncsa.uiuc.edu/nra/HDF5. 1998;
23. Rew R, Davis G. NetCDF: An interface for scientific data access. Computer Graphics and Applications, IEEE. IEEE; 1990; 10: 76–82. doi: 10.1109/38.56302
24. Wickham H. Tidy Data. Journal of Statistical Software. 2014; 59: 1–23. http://www.jstatsoft.org/v59/i10 doi: 10.18637/jss.v059.i10
25. Buffalo V. Bioinformatics data skills: Reproducible and robust research with open source tools. O'Reilly Media, Inc. 2015.
26. NCBI. NCBI is phasing out sequence gIs—use accession.Version instead! http://www.ncbi.nlm.nih.gov/news/03-02-2016-phase-out-of-GI-numbers/;
27. Preston-Werner T. Semantic Versioning 2.0.0. http://semver.org; 2014.
28. Johnson GT, Goodsell DS, Autin L, Forli S, Sanner MF, Olson AJ. 3D molecular models of whole HIV-1 virions generated with cellPACK. Faraday Discuss. The Royal Society of Chemistry; 2014; 169: 23–44. doi: 10.1039/C4FD00017J PMID: 25253262
29. Strasser C, Cook R, Michener W, Budden A. Primer on Data Management: What you always wanted to know [Internet]. California Digital Libraries; 2012. doi: 10.5060/D2251G48
30. Goodin D. Poorly anonymized logs reveal NYC cab drivers' detailed whereabouts [Internet]. 2015. http://arstechnica.com/tech-policy/2014/06/poorly-anonymized-logs-reveal-nyc-cab-drivers-detailed-whereabouts/
31. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS genetics. Public Library of Science; 2008; 4: e1000167. doi: 10.1371/journal.pgen.1000167 PMID: 18769715
32. Kahn SD. On the future of genomic data. Science (New York, NY). 2011; 331: 728–9. doi: 10.1126/science.1197891 PMID: 21311016
33. Wandelt S, Bux M, Leser U. Trends in genome compression. Current Bioinformatics. Bentham Science Publishers; 2014; 9: 315–326. doi: 10.2174/1574893609666140516010143
34. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. International journal of epidemiology. 2014; 43: 1929–44. doi: 10.1093/ije/dyu188 PMID: 25261970
35. Briney K. Rule of 3. DATA AB INITIO. http://dataabinitio.com/?p=320; 2013.
36. Henkel H, Hutchison V, Strasser C, Rebich Hespanha S, Vanderbilt K, Wayne L, et al. DataONE Education Modules. https://www.dataone.org/education-modules; 2012.
37. Witt M, Carlson J, Brandt S, Cragin M. Constructing Data Curation Profiles. International Journal of Digital Curation. 2009; 4: 93–103. doi: 10.2218/ijdc.v4i3.117

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Effective Statistical Practice

**Robert E. Kass[1], Brian S. Caffo[2], Marie Davidian[3], Xiao-Li Meng[4], Bin Yu[5], Nancy Reid[6]***

**1** Department of Statistics, Machine Learning Department, and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America, **2** Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, United States of America, **3** Department of Statistics, North Carolina State University, Raleigh, North Carolina, United States of America, **4** Department of Statistics, Harvard University, Cambridge, Massachusetts, United States of America, **5** Department of Statistics and Department of Electrical Engineering and Computer Science, University of California Berkeley, Berkeley, California, United States of America, **6** Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

* reid@utstat.utoronto.ca





**Citation:** Kass RE, Caffo BS, Davidian M, Meng X-L, Yu B, Reid N (2016) Ten Simple Rules for Effective Statistical Practice. PLoS Comput Biol 12(6): e1004961. doi:10.1371/journal.pcbi.1004961

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** June 9, 2016



**Funding:** BSC's research is partially supported by the National Institutes of Health grant EB012547: www.nibib.nih.gov. MD's research is partially supported by the National Institutes of Health grant NIH P01 CA142538: www.nih.gov. REK's research is partially supported by the National Institute of Mental Health R01 MH064537: www.nimh.nih.gov. NR's research is partially supported by the Natural Sciences and Engineering Council of Canada grant RGPIN-2015-06390: www.nserc.ca. BY's research is partially supported by the National Science Foundation grant CCF-0939370: www.nsf.gov. XLM's research is partially supported by the National Science Foundation (www.nsf.gov) DMS 1513492. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

Several months ago, Phil Bourne, the initiator and frequent author of the wildly successful and incredibly useful "Ten Simple Rules" series, suggested that some statisticians put together a Ten Simple Rules article related to statistics. (One of the rules for writing a PLOS Ten Simple Rules article is to be Phil Bourne [1]. In lieu of that, we hope effusive praise for Phil will suffice.)

Implicit in the guidelines for writing Ten Simple Rules [1] is "know your audience." We developed our list of rules with researchers in mind: researchers having some knowledge of statistics, possibly with one or more statisticians available in their building, or possibly with a healthy do-it-yourself attitude and a handful of statistical packages on their laptops. We drew on our experience in both collaborative research and teaching, and, it must be said, from our frustration at being asked, more than once, to "take a quick look at my student's thesis/my grant application/my referee's report: it needs some input on the stats, but it should be pretty straightforward."

There are some outstanding resources available that explain many of these concepts clearly and in much more detail than we have been able to do here: among our favorites are Cox and Donnelly [2], Leek [3], Peng [4], Kass et al. [5], Tukey [6], and Yu [7].

Every article on statistics requires at least one caveat. Here is ours: we refer in this article to "science" as a convenient shorthand for investigations using data to study questions of interest. This includes social science, engineering, digital humanities, finance, and so on. Statisticians are not shy about reminding administrators that statistical science has an impact on nearly every part of almost all organizations.

## Rule 1: Statistical Methods Should Enable Data to Answer Scientific Questions

A big difference between inexperienced users of statistics and expert statisticians appears as soon as they contemplate the uses of some data. While it is obvious that experiments generate data to answer scientific questions, inexperienced users of statistics tend to take for granted the link between data and scientific issues and, as a result, may jump directly to a technique based on data structure rather than scientific goal. For example, if the data were in

**Competing Interests:** The authors have declared that no competing interests exist.

a table, as for microarray gene expression data, they might look for a method by asking, "Which test should I use?" while a more experienced person would, instead, start with the underlying question, such as, "Where are the differentiated genes?" and, from there, would consider multiple ways the data might provide answers. Perhaps a formal statistical test would be useful, but other approaches might be applied as alternatives, such as heat maps or clustering techniques. Similarly, in neuroimaging, understanding brain activity under various experimental conditions is the main goal; illustrating this with nice images is secondary. This shift in perspective from statistical technique to scientific question may change the way one approaches data collection and analysis. After learning about the questions, statistical experts discuss with their scientific collaborators the ways that data might answer these questions and, thus, what kinds of studies might be most useful. Together, they try to identify potential sources of variability and what hidden realities could break the hypothesized links between data and scientific inferences; only then do they develop analytic goals and strategies. This is a major reason why collaborating with statisticians can be helpful, and also why the collaborative process works best when initiated early in an investigation. See Rule 3.

## Rule 2: Signals Always Come with Noise

Grappling with variability is central to the discipline of statistics. Variability comes in many forms. In some cases variability is good, because we need variability in predictors to explain variability in outcomes. For example, to determine if smoking is associated with lung cancer, we need variability in smoking habits; to find genetic associations with diseases, we need genetic variation. Other times variability may be annoying, such as when we get three different numbers when measuring the same thing three times. This latter variability is usually called "noise," in the sense that it is either not understood or thought to be irrelevant. Statistical analyses aim to assess the signal provided by the data, the interesting variability, in the presence of noise, or irrelevant variability.

A starting point for many statistical procedures is to introduce a mathematical abstraction: outcomes, such as patients being diagnosed with specific diseases or receiving numerical scores on diagnostic tests, will vary across the set of individuals being studied, and statistical formalism describes such variation using probability distributions. Thus, for example, a data histogram might be replaced, in theory, by a probability distribution, thereby shifting attention from the raw data to the numerical parameters that determine the precise features of the probability distribution, such as its shape, its spread, or the location of its center. Probability distributions are used in statistical models, with the model specifying the way signal and noise get combined in producing the data we observe, or would like to observe. This fundamental step makes statistical inferences possible. Without it, every data value would be considered unique, and we would be left trying to figure out all the detailed processes that might cause an instrument to give different values when measuring the same thing several times. Conceptualizing signal and noise in terms of probability within statistical models has proven to be an extremely effective simplification, allowing us to capture the variability in data in order to express uncertainty about quantities we are trying to understand. The formalism can also help by directing us to look for likely sources of systematic error, known as bias.

Big data makes these issues more important, not less. For example, Google Flu Trends debuted to great excitement in 2008, but turned out to overestimate the prevalence of influenza by nearly 50%, largely due to bias caused by the way the data were collected; see Harford [8], for example.

## Rule 3: Plan Ahead, Really Ahead

When substantial effort will be involved in collecting data, statistical issues may not be captured in an isolated statistical question such as, "What should my *n* be?" As we suggested in Rule 1, rather than focusing on a specific detail in the design of the experiment, someone with a lot of statistical experience is likely to step back and consider many aspects of data collection in the context of overall goals and may start by asking, "What would be the ideal outcome of your experiment, and how would you interpret it?" In trying to determine whether observations of X and Y tend to vary together, as opposed to independently, key issues would involve the way X and Y are measured, the extent to which the measurements represent the underlying conceptual meanings of X and Y, the many factors that could affect the measurements, the ability to control those factors, and whether some of those factors might introduce systematic errors (bias).

In Rule 2 we pointed out that statistical models help link data to goals by shifting attention to theoretical quantities of interest. For example, in making electrophysiological measurements from a pair of neurons, a neurobiologist may take for granted a particular measurement methodology along with the supposition that these two neurons will represent a whole class of similar neurons under similar experimental conditions. On the other hand, a statistician will immediately wonder how the specific measurements get at the issue of co-variation; what the major influences on the measurements are, and whether some of them can be eliminated by clever experimental design; what causes variation among repeated measurements, and how quantitative knowledge about sources of variation might influence data collection; and whether these neurons may be considered to be sampled from a well-defined population, and how the process of picking that pair could influence subsequent statistical analyses. A conversation that covers such basic issues may reveal possibilities an experimenter has not yet considered.

Asking questions at the design stage can save headaches at the analysis stage: careful data collection can greatly simplify analysis and make it more rigorous. Or, as Sir Ronald Fisher put it: "To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of" [9]. As a good starting point for reading on planning of investigations, see Chapters 1 through 4 of [2].

## Rule 4: Worry about Data Quality

Well-trained experimenters understand instinctively that, when it comes to data analysis, "garbage in produces garbage out." However, the complexity of modern data collection requires many assumptions about the function of technology, often including data pre-processing technology. It is highly advisable to approach pre-processing with care, as it can have profound effects that easily go unnoticed.

Even with pre-processed data, further considerable effort may be needed prior to analysis; this is variously called "data cleaning," "data munging," or "data carpentry." Hands-on experience can be extremely useful, as data cleaning often reveals important concerns about data quality, in the best case confirming that what was measured is indeed what was intended to be measured and, in the worst case, ensuring that losses are cut early.

Units of measurement should be understood and recorded consistently. It is important that missing data values can be recognized as such by relevant software. For example, 999 may signify the number 999, or it could be code for "we have no clue." There should be a defensible rule for handling situations such as "non-detects," and data should be scanned for anomalies such as variable 27 having half its values equal to 0.00027. Try to understand as much as you can how these data arrived at your desk or disk. Why are some data missing or incomplete? Did they get lost through some substantively relevant mechanism? Understanding such mechanisms can help to avoid some seriously misleading results. For example, in a developmental imaging study of attention deficit hyperactivity disorder, might some data have been lost from children with the most severe hyperactivity because they could not sit still in the MR scanner?

Once the data have been wrestled into a convenient format, have a look! Tinkering around with the data, also known as exploratory data analysis, is often the most informative part of the analysis. Exploratory plots can reveal data quality issues and outliers. Simple summaries, such as means, standard deviations, and quantiles, can help refine thinking and offer face validity checks for hypotheses. Many studies, especially when going in completely new scientific directions, are exploratory by design; the area may be too novel to include clear a priori hypotheses. Working with the data informally can help generate new hypotheses and ideas. However, it is also important to acknowledge the specific ways data are selected prior to formal analyses and to consider how such selection might affect conclusions. And it is important to remember that using a single set of data to both generate and test hypotheses is problematic. See Rule 9.

## Rule 5: Statistical Analysis Is More Than a Set of Computations

Statistical software provides tools to assist analyses, not define them. The scientific context is critical, and the key to principled statistical analysis is to bring analytic methods into close correspondence with scientific questions. See Rule 1. While it can be helpful to include references to a specific algorithm or piece of software in the Methods section of a paper, this should not be a substitute for an explanation of the choice of statistical method in answering a question. A reader will likely want to consider the fundamental issue of whether the analytic technique is appropriately linked to the substantive questions being answered. Don't make the reader puzzle over this: spell it out clearly.

At the same time, a structured algorithmic approach to the steps in your analysis can be very helpful in making this analysis reproducible by yourself at a later time, or by others with the same or similar data. See Rule 10.

## Rule 6: Keep it Simple

All else being equal, simplicity trumps complexity. This rule has been rediscovered and enshrined in operating procedures across many domains and variously described as "Occam's razor," "KISS," "less is more," and "simplicity is the ultimate sophistication." The principle of parsimony can be a trusted guide: start with simple approaches and only add complexity as needed, and then only add as little as seems essential.

Having said this, scientific data have detailed structure, and simple models can't always accommodate important intricacies. The common assumption of independence is often incorrect and nearly always needs careful examination. See Rule 8. Large numbers of measurements, interactions among explanatory variables, nonlinear mechanisms of action, missing data, confounding, sampling biases, and so on, can all require an increase in model complexity.

Keep in mind that good design, implemented well, can often allow simple methods of analysis to produce strong results. See Rule 3. Simple models help us to create order out of complex phenomena, and simple models are well suited for communication to our colleagues and the wider world.

## Rule 7: Provide Assessments of Variability

Nearly all biological measurements, when repeated, exhibit substantial variation, and this creates uncertainty in the result of every calculation based on the data. A basic purpose of statistical analysis is to help assess uncertainty, often in the form of a standard error or confidence interval, and one of the great successes of statistical modeling and inference is that it can provide estimates of standard errors from the same data that produce estimates of the quantity of interest. When reporting results, it is essential to supply some notion of statistical uncertainty. A common mistake is to calculate standard errors without taking into account the dependencies among data or variables, which usually means a substantial underestimate of the real uncertainty. See Rule 8.

Remember that every number obtained from the data by some computation would change somewhat, even if the measurements were repeated on the same biological material. If you are using new material, you can add to the measurement variability an increase due to the natural variability among samples. If you are collecting data on a different day, in a different lab, or under a slightly changed protocol, there are now three more potential sources of variability to be accounted for. In microarray analysis, batch effects are well known to introduce extra variability, and several methods are available to filter these. Extra variability means extra uncertainty in the conclusions, and this uncertainty needs to be reported. Such reporting is invaluable for planning the next investigation.

It is a very common feature of big data that uncertainty assessments tend to be overly optimistic (Cox [10], Meng [11]). For an instructive, and beguilingly simple, quantitative analysis most relevant to surveys, see the "data defect" section of [11]. Big data is not always as big as it looks: a large number of measurements on a small number of samples requires very careful estimation of the standard error, not least because these measurements are quite likely to be dependent.

## Rule 8: Check Your Assumptions

Every statistical inference involves assumptions, which are based on substantive knowledge and some probabilistic representation of data variation—this is what we call a statistical model. Even the so-called "model-free" techniques require assumptions, albeit less restrictive assumptions, so this terminology is somewhat misleading.

The most common statistical methods involve an assumption of linear relationships. For example, the ordinary correlation coefficient, also called the Pearson correlation, is a measure of linear association. Linearity often works well as a first approximation or as a depiction of a general trend, especially when the amount of noise in the data makes it difficult to distinguish between linear and nonlinear relationships. However, for any given set of data, the appropriateness of the linear model is an empirical issue and should be investigated.

In many ways, a more worrisome, and very common, assumption in statistical analysis is that multiple observations in the data are statistically independent. This is worrisome because relatively small deviations from this assumption can have drastic effects. When measurements are made across time, for example, the temporal sequencing may be important; if it is, specialized methods appropriate for time series need to be considered.

In addition to nonlinearity and statistical dependence, missing data, systematic biases in measurements, and a variety of other factors can cause violations of statistical modeling assumptions, even in the best experiments. Widely available statistical software makes it easy to perform analyses without careful attention to inherent assumptions, and this risks inaccurate, or even misleading, results. It is therefore important to understand the assumptions embodied in the methods you are using and to do whatever you can to understand and assess those assumptions. At a minimum, you will want to check how well your statistical model fits the data. Visual displays and plots of data and residuals from fitting are helpful for evaluating the relevance of assumptions and the fit of the model, and some basic techniques for assessing model fit are available in most statistical software. Remember, though, that several models can "pass the fit test" on the same data. See Rule 1 and Rule 6.

## Rule 9: When Possible, Replicate!

Every good analyst examines the data at great length, looking for patterns of many types and searching for predicted and unpredicted results. This process often involves dozens of procedures, including many alternative visualizations and a host of numerical slices through the data. Eventually, some particular features of the data are deemed interesting and important, and these are often the results reported in the resulting publication.

When statistical inferences, such as *p*-values, follow extensive looks at the data, they no longer have their usual interpretation. Ignoring this reality is dishonest: it is like painting a bull's eye around the landing spot of your arrow. This is known in some circles as *p*-hacking, and much has been written about its perils and pitfalls: see, for example, [12] and [13].

Recently there has been a great deal of criticism of the use of *p*-values in science, largely related to the misperception that results can't be worthy of publication unless "$p$ is less than 0.05." The recent statement from the American Statistical Association (ASA) [14] presents a detailed view of the merits and limitations of the *p*-value.

Statisticians tend to be aware of the most obvious kinds of data snooping, such as choosing particular variables for a reported analysis, and there are methods that can help adjust results in these cases; the False Discovery Rate method of Benjamini and Hochberg [15] is the basis for several of these.

For some analyses, there may be a case that some kinds of preliminary data manipulation are likely to be innocuous. In other situations, analysts may build into their work an informal check by trusting only extremely small *p*-values. For example, in high energy physics, the requirement of a "5-sigma" result is at least partly an approximate correction for what is called the "look-elsewhere effect."

The only truly reliable solution to the problem posed by data snooping is to record the statistical inference procedures that produced the key results, together with the features of the data to which they were applied, and then to replicate the same analysis using new data. Independent replications of this type often go a step further by introducing modifications to the experimental protocol, so that the replication will also provide some degree of robustness to experimental details.

Ideally, replication is performed by an independent investigator. The scientific results that stand the test of time are those that get confirmed across a variety of different, but closely related, situations. In the absence of experimental replications, appropriate forms of data perturbation can be helpful (Yu [16]). In many contexts, complete replication is very difficult or impossible, as in large-scale experiments such as multi-center clinical trials. In such cases, a minimum standard would be to follow Rule 10.

## Rule 10: Make Your Analysis Reproducible

In our current framework for publication of scientific results, the independent replication discussed in Rule 9 is not practical for most investigators. A different standard, which is easier to achieve, is reproducibility: given the same set of data, together with a complete description of the analysis, it should be possible to reproduce the tables, figures, and statistical inferences. However, even this lower standard can face multiple barriers, such as different computing architectures, software versions, and settings.

One can dramatically improve the ability to reproduce findings by being very systematic about the steps in the analysis (see Rule 5), by sharing the data and code used to produce the results, and by following Goodman et al. [17]. Modern reproducible research tools like Sweave [18], knitr [19], and iPython [20] notebooks take this a step further and combine the research report with the code. Reproducible research is itself an ongoing area of research and a very important area that we all need to pay attention to.

## Conclusion

Mark Twain popularized the saying, "There are three kinds of lies: lies, damned lies, and statistics." It is true that data are frequently used selectively to give arguments a false sense of support. Knowingly misusing data or concealing important information about the way data and data summaries have been obtained is, of course, highly unethical. More insidious, however, are the widespread instances of claims made about scientific hypotheses based on well-intentioned yet faulty statistical reasoning. One of our chief aims here has been to emphasize succinctly many of the origins of such problems and ways to avoid the pitfalls.

A central and common task for us as research investigators is to decipher what our data are able to say about the problems we are trying to solve. Statistics is a language constructed to assist this process, with probability as its grammar. While rudimentary conversations are possible without good command of the language (and are conducted routinely), principled statistical analysis is critical in grappling with many subtle phenomena to ensure that nothing serious will be lost in translation and to increase the likelihood that your research findings will stand the test of time. To achieve full fluency in this mathematically sophisticated language requires years of training and practice, but we hope the Ten Simple Rules laid out here will provide some essential guidelines.

Among the many articles reporting on the ASA's statement on *p*-values, we particularly liked a quote from biostatistician Andrew Vickers in [21]: "Treat statistics as a science, not a recipe." This is a great candidate for Rule 0.

## Acknowledgments

We consulted many colleagues informally about this article, but the opinions expressed here are unique to our small committee of authors. We'd like to give a shout out to xkcd.com for conveying statistical ideas with humor, to the Simply Statistics blog as a reliable source for thoughtful commentary, to FiveThirtyEight for bringing statistics to the world (or at least to the media), to Phil Bourne for suggesting that we put together this article, and to Steve Pierson of the American Statistical Association for getting the effort started.

## References

1. Dashnow H, Lonsdale A, Bourne PE (2014) Ten simple rules for writing a PLOS ten simple rules article. PLoS Comput Biol 10(10): e1003858. doi: 10.1371/journal.pcbi.1003858 PMID: 25340653
2. Cox DR, Donnelly CA (2011) Principles of Applied Statistics. Cambridge: Cambridge University Press.
3. Leek JT (2015) The Elements of Data Analytic Style. Leanpub, https://leanpub.com/artofdatascience.
4. Peng R (2014) The Art of Data Science. Leanpub, https://leanpub.com/artofdatascience.
5. Kass RE, Eden UT, Brown EN (2014) Analysis of Neural Data. Springer: New York.
6. Tukey JW (1962) The future of data analysis. Ann Math Stat 33: 1–67.
7. Yu B (2013) Stability. Bernoulli, 19(4): 1484–1500.
8. Harford T (2015) Big Data: are we making a big mistake? Significance 11: 14–19.
9. Fisher RA (1938) Presidential address. Sankhyā 4: 14–17.
10. Cox DR (2015) Big data and precision. Biometrika 102: 712–716.

11. Meng XL (2014) A trio of inference problems that could win you a Nobel prize in statistics (if you help fund it). In: Lin X, Genest C, Banks DL, Molenberghs G, Scott DW, Wang J-L,editors. Past, Present, and Future of Statistical Science, Boca Raton: CRC Press. pp. 537–562.
12. Gelman A, Loken E (2014) The statistical crisis in science. Am Sci 102: 460–465
13. Aschwanden C (2015) Science isn't broken. August 11 2015 http://fivethirtyeight.com/features/science-isnt-broken/
14. Wasserstein RL, Lazar NA (2016) The ASA's statement on p-values: context, process, and purpose, The American Statistician doi: 10.1080/00031305.2016.1154108
15. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57: 289–300.
16. Yu, B (2015) Data wisdom for data science. April 13 2015 http://www.odbms.org/2015/04/data-wisdom-for-data-science/
17. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, et al. (2014) Ten simple rules for the care and feeding of scientific data. PLoS Comput Biol 10(4): e1003542. doi: 10.1371/journal.pcbi.1003858 PMID: 24763340
18. Leisch F (2002) Sweave: Dynamic generation of statistical reports using data analysis. In Härdle W, Rönz H, editors. Compstat: Proceedings in Computational Statistics, Heidelberg: Springer-Verlag, pp. 575–580.
19. Xie Y (2014) Dynamic Documents with R and knitr. Boca Raton: CRC Press.
20. Pérez F, Granger BE (2007) IPython: A system for interactive scientific computing. Comput Sci Eng 9 (3), 21–29.
21. Baker M (2016) Statisticians issue warning over misuse of *P* values. Nature 531, (151) doi: 10.1038/nature.2016.19503

EDITORIAL

# Ten Simple Rules for Selecting a Bio-ontology

**James Malone[1]*, Robert Stevens[2], Simon Jupp[1], Tom Hancocks[1], Helen Parkinson[1], Cath Brooksbank[1]**

**1** European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, **2** School of Computer Science, University of Manchester, Manchester, United Kingdom

* james@factbio.com

CrossMark click for updates

OPEN ACCESS

**Citation:** Malone J, Stevens R, Jupp S, Hancocks T, Parkinson H, Brooksbank C (2016) Ten Simple Rules for Selecting a Bio-ontology. PLoS Comput Biol 12(2): e1004743. doi:10.1371/journal.pcbi.1004743

**Editor:** Scott Markel, UNITED STATES

**Published:** February 11, 2016



**Funding:** EMBL and European Union grants Diachron (601043) and BioMedBridges (284209). National Institutes of Health NCBC grant for National Center for Biomedical Ontology [U54-HG004028]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Biologists and bioinformaticians now look to ontologies or software that uses ontologies as a means of standardising the way data are described, queried, and interpreted. Ontologies can be used for the annotation and curation of experimental datasets and, in data sharing, both within and beyond the confines of individual labs, organizations, and communities. Bio-ontologies are also commonly used in methods of analysis, particularly in gene set enrichment analysis [1], using ontologies such as the Gene Ontology. With modern high-throughput data-generation technologies, there is now, more than ever, a need to integrate data from these and other sources, and there is a concomitant need for ontologies—raising the question of how to choose a bio-ontology.

Over the past decade, a community has grown up around the success of efforts to harmonise the semantic description of biological entities, with ontologies exemplified in the emergence of the Open Biological and Biomedical Ontologies (OBO) Foundry [2]. These efforts were first led by the aforementioned Gene Ontology [3] and have expanded to ontologies that describe a significant range of the primary areas of biology and its science. Exploring bio-ontologies through browsers such as the Ontology Lookup Service [4] at the European Bioinformatics Institute and BioPortal [5] at the National Center for Biomedical Ontology (NCBO) —whose existence is itself a measure of the community size—shows there are over 400 ontologies containing, collectively, over 5 million classes (by classes, we mean ontological terms together with their associated descriptions and synonyms). These ontologies cover areas such as diseases [6], phenotypes [7], anatomy [8], experimental conditions [9,10], cell types [11], and bioinformatics software [12].

There are now many ontologies from which to choose, but which ontology should be chosen? In order to answer this question, we present ten simple rules that should help to guide the choice of a bio-ontology. The rules are designed to be useful for those wishing consume a bio-ontology. Users of bio-ontologies are varied in their profile and include data curators, application developers, and, of course, ontology developers who may be consuming part of an ontology in their own work.

## Rule 1: The Ontology Should Be about a Specific Domain of Knowledge

Specifically, an ontology should provide coverage for the area it claims to describe. Although almost no ontology is complete, you should aim to find an ontology that describes a considerable amount of the area to which it lays claim. It should also describe the field of interest in such a way that extensions to cover missing areas are possible without a major rewriting of the ontology. Missing terms are to be expected, but if the ontology is missing large areas that are key to describing the domain, then it may not be a suitable ontology. For instance, a disease ontology that does not include cancers would be inadequate if the diseases that you were aiming to describe included cancers. Furthermore, even if you personally don't have any cancer data to describe, you need to consider the notion that a disease ontology with such a large gap in it is unlikely to gain wide community adoption. Conversely, if an ontology claims to only describe a specific subset of a domain, or even just a local application, then it should do so appropriately and should not be considered an unsuitable ontology simply because it has a more limited scope; it may be less useful to the wider world, but this does not make it a bad ontology. One computational service that can help a user estimate whether or not an ontology can provide coverage is the NCBO Ontology Recommender [13]. Recommender measures whether an ontology from the NCBO BioPortal matches a given set of text based on a measure of coverage, which can help to inform whether a given ontology contains the terms a user might be expecting.

## Rule 2: The Ontology Should Reflect Current Understanding of Biological Systems

Unless the aim of the ontology is to capture a historic viewpoint (a legitimate objective), then it should reflect current science, or at least not contradict it. For instance, the old dogma of DNA to RNA without feedback would no longer be accepted in modern biology. It is better to make statements that are too broad but remain correct rather than make specific statements that are wrong. The correctness of an ontology is often evaluated using techniques such as competency questions. [14]. Competency questions are queries that are required to be answerable by an ontology, with the returned answers thus demonstrating whether or not an ontology is giving correct, expected answers. In the simplest form, a user may ask for subtypes of a class, e.g., "What are the subclasses of fat cell?", but this can also be more complex depending upon needs, e.g., "Which cell lines are derived from human, epithelial cells and are taken from melanoma samples?" Correct answers suggest the ontology is reflecting current science correctly.

## Rule 3: The Ontology Classes and Relationships Should Persist

One of the primary use cases of an ontology is to describe biomedical data through annotations; disconnecting the descriptions of this data from their semantics through the deletion of the ontology identifiers undermines the advantage of using ontologies in the first instance. This is crucial if these ontologies' annotations are being used for data sharing, integration, or analysis. Identifiers in most biomedical ontologies are formed using accessioned IDs rather than textual labels, with the intent of removing potential ID clashes and decoupling the textual part of a class (i.e., the label) from the identifier referring to it. This has the advantage of enabling small modifications to a class label without affecting the class identifier, where the class is still referring to the same entity. In cases in which the identifier is a Uniform Resource Identifier (URI), these should resolve to provide both human-readable and machine-interpretable information. Services like Identifiers.org provide a URI resolution service for many biomedical ontologies. Identifiers should be maintained, and if it is necessary to remove a class, it should be labelled as "obsolete" rather than simply deleted. Maintaining this audit trail is the sign of a well-managed ontology—deleting identifiers is the sign of a poorly managed ontology.

## Rule 4: Classes Should Contain Textual Definitions

This is crucial for users who come to an ontology trying to understand what a particular class is describing. It may, on occasion, be obvious—"*Homo sapiens*," for instance. On others, it is critical—a cell line named "Bas666" could be difficult to interpret. An additional sign of a suitable ontology is that it contains appropriate synonyms (e.g., "human" for "*Homo sapiens*") and related alternative terms (e.g., in Gene Ontology, use of narrow synonyms such as "type I programmed cell death" for the label "apoptotic process"), since language can differ between communities and specialities even though the underlying class being described is the same thing. As well as textual definitions, many ontologies also contain logical descriptions of the class that are amenable to computational interpretation. These descriptions use rules, or "axioms," to relate a class to other classes, such as describing that a heart is part of the cardiovascular system. Whether or not such computational aspects are necessary for a particular use case should form part of the decision when selecting a bio-ontology.

## Rule 5: Textual Definitions Should Be Written for Domain Experts

Creators of ontologies often fall into the trap of defining classes using ontology jargon (often philosophical in nature). This may make them understandable to ontologists and/or philosophers, but this is not useful if the language used means nothing to the ontology's user community. A good ontology will reflect commonly used nomenclature in naming classes within it. Similarly, textual definitions should also reflect common language used in the biological domain. Textual definitions and labels that include ontology jargon are the sign of an unsuitable ontology. Ontologists should accurately describe the biomedical domain without modifying it.

## Rule 6: The Ontology Should Be Developed by the Community but Not Incapacitated by It

Reflecting current science is a difficult task, given the growing knowledge of the breadth and depth of entities in the domain. Gaining community consensus is a noble cause and should help to reflect current science correctly and enhance opportunities for wide adoption of an ontology. It is also almost always better to work towards getting entities added into an existing ontology that is supported by a community rather than inventing a new ontology. Engaging with the community, however, should not deflect from the task of developing an ontology. Decision making—should a user ask for a new class, for example—should not take months while consensus is obtained. Similarly, a lone gatekeeper making all the decisions about what happens within an ontology is also a bad sign. Most ontologies will have a public forum for dealing with user requests, and looking at mailing list archives or issue trackers (e.g., Gene Ontology http://geneontology.org/page/go-mailing-lists) will provide insight on how the ontology is being developed.

Another aspect of collaborating is that of compromise. Typically, everyone has an opinion on the science in which they are interested, and typically they don't all align, so there is an element of compromise to selecting an ontology. A favoured label might not exist in the ontology, but rejecting wholesale a community-developed ontology in favour of inventing a de novo artefact with one's own favourite labels is not always the best option. As above though, there are circumstances under which this may be ultimately the better option; here, the balance is in weighing up requirements and making a judgement based on what is most important. Wider integration with a community is a good thing when it works.

## Rule 7: The Ontology Should Be under Active Development

An ontology should have a dedicated presence on the web, such as a project website that provides information on how to contact the developers and contribute to the ontology. Any associated mailing lists or version control systems can be used to gauge recent activity on the ontology. Recent work [15] has shown that it is possible to describe how actively an ontology is maintained and in what way it is being modified. In general, an ontology that is not actively developed and has not been updated for many months or years is unlikely to respond to new requirements should they occur.

## Rule 8: Previous Versions Should Be Available

Given the changing nature of data and, hopefully, of the ontology as it updates to reflect current knowledge, data annotations can become out of date. An ontology should provide a clear versioning and release policy. It is important to be able to access older versions of an ontology so the context of data descriptions can be understood. This also relates to Rule 3 about not deleting ontology classes, and in turn both rules relate to enabling reproducibility of data analysis. Being unable to trace provenance of data annotations made with an older version of an ontology is a barrier to future reproducibility; selecting an ontology that maintains previous versions is therefore an important consideration.

## Rule 9: Open Data Requires Open Ontologies

An important consideration is whether or not the ontology is being selected for use with open data with the intention of wider sharing. Using an ontology that is restrictive in licensing can also have an impact on the data described with such an ontology, restricting access to the semantics which are necessary to understand it. If data sharing is the aim, then using ontologies with permissive licenses should be a priority. Permissive licenses, such as those developed by Creative Commons (https://creativecommons.org/licenses/), can be used to communicate both how the work (ontology) can be exploited and how attribution should be given. One of the important outcomes of the standardization efforts from the OBO Foundry has been the widespread use of the OBO Relation Ontology (RO), an open ontology of biomedical relationships. The widespread use of RO has led to a de facto standard in much of the bio-ontology world, which has had positive implications on integration of resources, facilitated by the open license with which the RO is released.

## Rule 10: Sometimes an Ontology Is Not Needed at All

Ontologies provide a means of "knowing" what is being described in a data set. There is, however, more than one way to capture such knowledge. Before embarking on using or indeed making a bio-ontology, you need to decide whether an ontology is really what is needed. In the broadest terms, we are talking about knowledge organisation systems of which there are numerous types of useful resources: glossaries, taxonomies, thesauri, ontologies, and terminologies. As a growing discipline, there is a temptation to suggest that using biomedical ontologies will offer some advantage. Ontologies offer advantages over other knowledge systems—they enable both computational use and human understanding, they can contain multiple classification axes of classes as well as formal descriptions of how classes relate to one another, and can include rich vocabularies of labels, synonyms, and textual definitions. If these are desirable selection criteria, then an ontology should be considered. Ontologies do also come with computational overheads, however, and can be complex to understand. Languages such as the Web Ontology Language (OWL) [16] utilise description logics, which are technically challenging. Other resources such as a vocabulary do not offer the sorts of classification and rich computational descriptions of an ontology but are often much simpler to understand. Let your requirements guide you; ontologies are not a panacea—sometimes one isn't needed at all.

## Conclusion

Bio-ontologies represent an important tool for describing metadata, an increasingly important consideration as the scientific community aims for open, reusable data. It is perhaps no surprise then that in the Ten Simple Rules for the Care and Feeding of Scientific Data [17], the word "metadata" appeared 11 times. The choice of which ontology to pick and even when to use one is not always straightforward, as demonstrated by the number of times the authors are asked to recommend a particular ontology for a given problem. The single most important consideration in selecting a bio-ontology is to understand requirements first before deciding to engage with a particular ontology or indeed before minting one's own ontology. By identifying needs and selecting current ontologies using the above rules, it is possible to reach a conclusion as to whether or not a resource is useful to a given user. Moreover, reusing ontologies that similarly satisfy another user's needs helps to spread the burden of development across the community and ensure we don't end up with islands of metadata, undermining the efforts of openness and sharing.

## Acknowledgments

We would like to thank all those involved in building open access bio-ontologies, especially those in the OBO community, and Judith Blake and Mark Musen for their comments on improving this article.

## References

1. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research. 2009; 37(1):1–13. http://nar.oxfordjournals.org/content/37/1/1.abstract. doi: 10.1093/nar/gkn923 PMID: 19033363
2. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology. 2007 November; 25(11):1251–1255. doi: 10.1038/nbt1346 PMID: 17989687
3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature genetics. 2000; 25(1):25–29. PMID: 10802651
4. Côté R, Reisinger F, Martens L, Barsnes H, Vizcaino JA, Hermjakob H. The Ontology Lookup Service: bigger and better. Nucleic Acids Research. 2010; 38(suppl 2):W155–W160. http://nar.oxfordjournals.org/content/38/suppl_2/W155.abstract.
5. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Research. 2009; 37(suppl 2):W170–W173.
6. Schriml LM, Arze C, Nadendla S, Chang YWW, Mazaitis M, Felix V, et al. Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Research. 2012; 40(D1):D940–D946. http://nar.oxfordjournals.org/content/40/D1/D940.abstract.
7. Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, et al. Phenotype ontologies: the bridge between genomics and evolution. Trends in Ecology and Evolution. 2007; 22(7):345–350. http://www.sciencedirect.com/science/article/pii/S0169534707001048. PMID: 17416439
8. Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, Haendel MA, et al. Uberon, an integrative multi-species anatomy ontology. Genome Biol. 2012; 13(1):R5. doi: 10.1186/gb-2012-13-1-r5 PMID: 22293552
9. Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, et al. Modeling sample variables with an Experimental Factor Ontology. Bioinformatics. 2010; 26(8):1112–1118. http://bioinformatics.oxfordjournals.org/content/26/8/1112.abstract. doi: 10.1093/bioinformatics/btq099 PMID: 20200009
10. Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, et al. Modeling biomedical experimental processes with OBI. J Biomed Semantics. 2010; 1(Suppl 1):S7. doi: 10.1186/2041-1480-1-S1-S7 PMID: 20626927
11. Bard J, Rhee SY, Ashburner M. An ontology for cell types. Genome biology. 2005; 6(2):R21. PMID: 15693950

12. Malone J, Brown A, Lister AL, Ison J, Hull D, Parkinson H, et al. The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation. Journal of Biomedical Semantics. 2014; 5(25).
13. Jonquet C, Musen M, Shah N. Building a biomedical ontology recommender web service. Journal of Biomedical Semantics. 2010; 1(S1).
14. Azzaoui K, Jacoby E, Senger S, Rodríguez EC, Loza M, Zdrazil B, et al. Scientific competency questions as the basis for semantically enriched open pharmacological space development. Drug Discovery Today. 2013; 18(17–18):843–852. doi: 10.1016/j.drudis.2013.05.008 PMID: 23702085
15. Malone J, Stevens R. Measuring the level of activity in community built bio-ontologies. Journal of Biomedical Informatics. 2013; 46(1):5–14. http://www.sciencedirect.com/science/article/pii/S153204641200055X. doi: 10.1016/j.jbi.2012.04.002 PMID: 22554701
16. OWL Working Group W. OWL 2 Web Ontology Language: Document Overview. W3C; 11 December 2012. http://www.w3.org/TR/owl2-overview/.
17. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol. 2014; 10(4):e1003542. doi: 10.1371/journal.pcbi.1003542 PMID: 24763340

PLOS COMPUTATIONAL BIOLOGY

PERSPECTIVE

# Ten Simple Rules for Creating a Good Data Management Plan

**William K. Michener***

College of University Libraries & Learning Sciences, University of New Mexico, Albuquerque, New Mexico, United States of America

* william.michener@gmail.com



OPEN ACCESS

**Citation:** Michener WK (2015) Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol 11(10): e1004525. doi:10.1371/journal.pcbi.1004525

**Editor:** Philip E. Bourne, National Institutes of Health, UNITED STATES

**Published:** October 22, 2015



**Funding:** This work was supported by NSF IIA-1301346, IIA-1329470, and ACI-1430508 (http://nsf.gov). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

## Introduction

Research papers and data products are key outcomes of the science enterprise. Governmental, nongovernmental, and private foundation sponsors of research are increasingly recognizing the value of research data. As a result, most funders now require that sufficiently detailed data management plans be submitted as part of a research proposal. A data management plan (DMP) is a document that describes how you will treat your data during a project and what happens with the data after the project ends. Such plans typically cover all or portions of the data life cycle—from data discovery, collection, and organization (e.g., spreadsheets, databases), through quality assurance/quality control, documentation (e.g., data types, laboratory methods) and use of the data, to data preservation and sharing with others (e.g., data policies and dissemination approaches). Fig 1 illustrates the relationship between hypothetical research and data life cycles and highlights the links to the rules presented in this paper. The DMP undergoes peer review and is used in part to evaluate a project's merit. Plans also document the data management activities associated with funded projects and may be revisited during performance reviews.

Earlier articles in the Ten Simple Rules series of *PLOS Computational Biology* provided guidance on getting grants [1], writing research papers [2], presenting research findings [3], and caring for scientific data [4]. Here, I present ten simple rules that can help guide the process of creating an effective plan for managing research data—the basis for the project's findings, research papers, and data products. I focus on the principles and practices that will result in a DMP that can be easily understood by others and put to use by your research team. Moreover, following the ten simple rules will help ensure that your data are safe and sharable and that your project maximizes the funder's return on investment.

## Rule 1: Determine the Research Sponsor Requirements

Research communities typically develop their own standard methods and approaches for managing and disseminating data. Likewise, research sponsors often have very specific DMP expectations. For instance, the Wellcome Trust, the Gordon and Betty Moore Foundation (GBMF), the United States National Institutes of Health (NIH), and the US National Science Foundation (NSF) all fund computational biology research but differ markedly in their DMP requirements. The GBMF, for instance, requires that potential grantees develop a comprehensive DMP in conjunction with their program officer that answers dozens of specific questions. In contrast, NIH requirements are much less detailed and primarily ask that potential grantees explain how data will be shared or provide reasons as to why the data cannot be shared. Furthermore, a

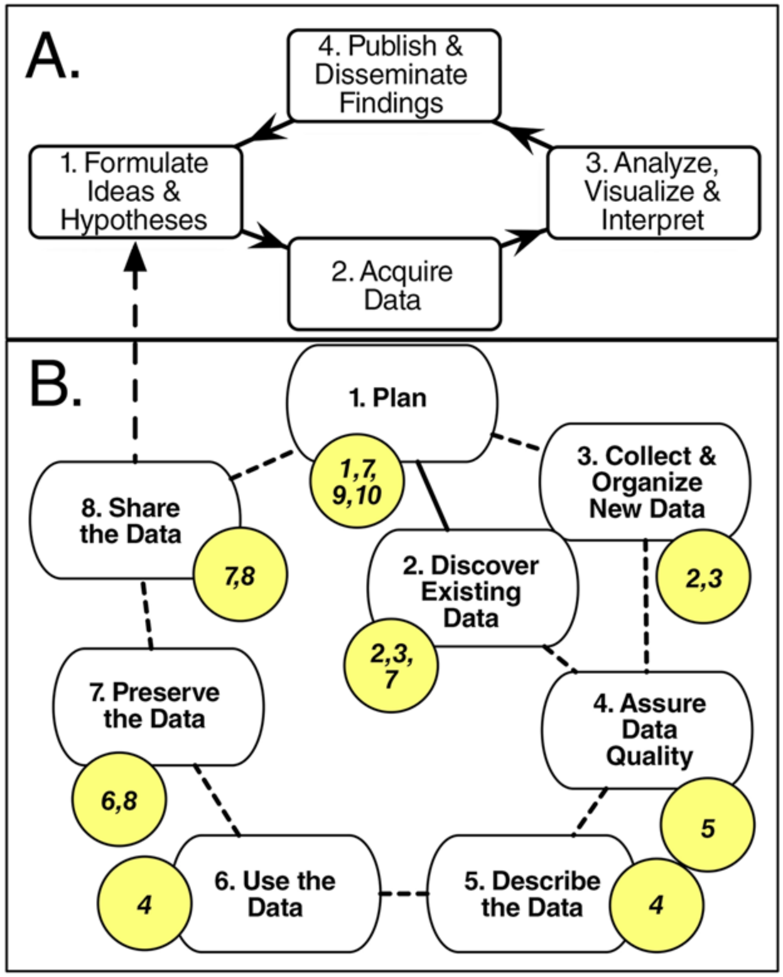



**Fig 1. Relationship of the research life cycle (A) to the data life cycle (B); note: highlighted circles refer to the rules that are most closely linked to the steps of the data life cycle.** As part of the research life cycle (A), many researchers (1) test ideas and hypotheses by (2) acquiring data that are (3) incorporated into various analyses and visualizations, leading to interpretations that are then (4) published in the literature and disseminated via other mechanisms (e.g., conference presentations, blogs, tweets), and that often lead back to (1) new ideas and hypotheses. During the data life cycle (B), researchers typically (1) develop a plan for how data will be managed during and after the project; (2) discover and acquire existing data and (3) collect and organize new data; (4) assure the quality of the data; (5) describe the data (i.e., ascribe metadata); (6) use the data in analyses, models, visualizations, etc.; and (7) preserve and (8) share the data with others (e.g., researchers, students, decision makers), possibly leading to new ideas and hypotheses.

doi:10.1371/journal.pcbi.1004525.g001

single research sponsor (such as the NSF) may have different requirements that are established for individual divisions and programs within the organization. Note that plan requirements may not be labeled as such; for example, the National Institutes of Health guidelines focus largely on data sharing and are found in a document entitled "NIH Data Sharing Policy and Implementation Guidance" (http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm).

Significant time and effort can be saved by first understanding the requirements set forth by the organization to which you are submitting a proposal. Research sponsors normally provide DMP requirements in either the public request for proposals (RFP) or in an online grant proposal guide. The DMPTool (https://dmptool.org/) and DMPonline (https://dmponline.dcc.ac.

uk/) websites are also extremely valuable resources that provide updated funding agency plan requirements (for the US and United Kingdom, respectively) in the form of templates that are usually accompanied with annotated advice for filling in the template. The DMPTool website also includes numerous example plans that have been published by DMPTool users. Such examples provide an indication of the depth and breadth of detail that are normally included in a plan and often lead to new ideas that can be incorporated in your plan.

Regardless of whether you have previously submitted proposals to a particular funding program, it is always important to check the latest RFP, as well as the research sponsor's website, to verify whether requirements have recently changed and how. Furthermore, don't hesitate to contact the responsible program officer(s) that are listed in a specific solicitation to discuss sponsor requirements or to address specific questions that arise as you are creating a DMP for your proposed project. Keep in mind that the principle objective should be to create a plan that will be useful for your project. Thus, good data management plans can and often do contain more information than is minimally required by the research sponsor. Note, though, that some sponsors constrain the length of DMPs (e.g., two-page limit); in such cases, a synopsis of your more comprehensive plan can be provided, and it may be permissible to include an appendix, supplementary file, or link.

## Rule 2: Identify the Data to Be Collected

Every component of the DMP depends upon knowing how much and what types of data will be collected. Data volume is clearly important, as it normally costs more in terms of infrastructure and personnel time to manage 10 terabytes of data than 10 megabytes. But, other characteristics of the data also affect costs as well as metadata, data quality assurance and preservation strategies, and even data policies. A good plan will include information that is sufficient to understand the nature of the data that is collected, including:

- **Types.** A good first step is to list the various types of data that you expect to collect or create. This may include text, spreadsheets, software and algorithms, models, images and movies, audio files, and patient records. Note that many research sponsors define data broadly to include physical collections, software and code, and curriculum materials.
- **Sources.** Data may come from direct human observation, laboratory and field instruments, experiments, simulations, and compilations of data from other studies. Reviewers and sponsors may be particularly interested in understanding if data are proprietary, are being compiled from other studies, pertain to human subjects, or are otherwise subject to restrictions in their use or redistribution.
- **Volume.** Both the total volume of data and the total number of files that are expected to be collected can affect all other data management activities.
- **Data and file formats.** Technology changes and formats that are acceptable today may soon be obsolete. Good choices include those formats that are nonproprietary, based upon open standards, and widely adopted and preferred by the scientific community (e.g., Comma Separated Values [CSV] over Excel [.xls, xlsx]). Data are more likely to be accessible for the long term if they are uncompressed, unencrypted, and stored using standard character encodings such as UTF-16.

The precise types, sources, volume, and formats of data may not be known beforehand, depending on the nature and uniqueness of the research. In such case, the solution is to iteratively update the plan (see Rule 9).

## Rule 3: Define How the Data Will Be Organized

Once there is an understanding of the volume and types of data to be collected, a next obvious step is to define how the data will be organized and managed. For many projects, a small number of data tables will be generated that can be effectively managed with commercial or open source spreadsheet programs like Excel and OpenOffice Calc. Larger data volumes and usage constraints may require the use of relational database management systems (RDBMS) for linked data tables like ORACLE or mySQL, or a Geographic Information System (GIS) for geospatial data layers like ArcGIS, GRASS, or QGIS.

The details about how the data will be organized and managed could fill many pages of text and, in fact, should be recorded as the project evolves. However, in drafting a DMP, it is most helpful to initially focus on the types and, possibly, names of the products that will be used. The software tools that are employed in a project should be amenable to the anticipated tasks. A spreadsheet program, for example, would be insufficient for a project in which terabytes of data are expected to be generated, and a sophisticated RDMBS may be overkill for a project in which only a few small data tables will be created. Furthermore, projects dependent upon a GIS or RDBMS may entail considerable software costs and design and programming effort that should be planned and budgeted for upfront (see Rules 9 and 10). Depending on sponsor requirements and space constraints, it may also be useful to specify conventions for file naming, persistent unique identifiers (e.g., Digital Object Identifiers [DOIs]), and versioning control (for both software and data products).

## Rule 4: Explain How the Data Will Be Documented

Rows and columns of numbers and characters have little to no meaning unless they are documented in some fashion. Metadata—the details about what, where, when, why, and how the data were collected, processed, and interpreted—provide the information that enables data and files to be discovered, used, and properly cited. Metadata include descriptions of how data and files are named, physically structured, and stored as well as details about the experiments, analytical methods, and research context. It is generally the case that the utility and longevity of data relate directly to how complete and comprehensive the metadata are. The amount of effort devoted to creating comprehensive metadata may vary substantially based on the complexity, types, and volume of data.

A sound documentation strategy can be based on three steps. First, identify the types of information that should be captured to enable a researcher like you to discover, access, interpret, use, and cite your data. Second, determine whether there is a community-based metadata schema or standard (i.e., preferred sets of metadata elements) that can be adopted. As examples, variations of the Dublin Core Metadata Initiative Abstract Model are used for many types of data and other resources, ISO (International Organization for Standardization) 19115 is used for geospatial data, ISA-Tab file format is used for experimental metadata, and Ecological Metadata Language (EML) is used for many types of environmental data. In many cases, a specific metadata content standard will be recommended by a target data repository, archive, or domain professional organization. Third, identify software tools that can be employed to create and manage metadata content (e.g., Metavist, Morpho). In lieu of existing tools, text files (e.g., readme.txt) that include the relevant metadata can be included as headers to the data files.

A best practice is to assign a responsible person to maintain an electronic lab notebook, in which all project details are maintained. The notebook should ideally be routinely reviewed and revised by another team member, as well as duplicated (see Rules 6 and 9). The metadata recorded in the notebook provide the basis for the metadata that will be associated with data products that are to be stored, reused, and shared.

## Rule 5: Describe How Data Quality Will Be Assured

Quality assurance and quality control (QA/QC) refer to the processes that are employed to measure, assess, and improve the quality of products (e.g., data, software, etc.). It may be necessary to follow specific QA/QC guidelines depending on the nature of a study and research sponsorship; such requirements, if they exist, are normally stated in the RFP. Regardless, it is good practice to describe the QA/QC measures that you plan to employ in your project. Such measures may encompass training activities, instrument calibration and verification tests, double-blind data entry, and statistical and visualization approaches to error detection. Simple graphical data exploration approaches (e.g., scatterplots, mapping) can be invaluable for detecting anomalies and errors.

## Rule 6: Present a Sound Data Storage and Preservation Strategy

A common mistake of inexperienced (and even many experienced) researchers is to assume that their personal computer and website will live forever. They fail to routinely duplicate their data during the course of the project and do not see the benefit of archiving data in a secure location for the long term. Inevitably, though, papers get lost, hard disks crash, URLs break, and tapes and other media degrade, with the result that the data become unavailable for use by both the originators and others. Thus, data storage and preservation are central to any good data management plan. Give careful consideration to three questions:

1. How long will the data be accessible?
2. How will data be stored and protected over the duration of the project?
3. How will data be preserved and made available for future use?

The answer to the first question depends on several factors. First, determine whether the research sponsor or your home institution have any specific requirements. Usually, all data do not need to be retained, and those that do need not be retained forever. Second, consider the intrinsic value of the data. Observations of phenomena that cannot be repeated (e.g., astronomical and environmental events) may need to be stored indefinitely. Data from easily repeatable experiments may only need to be stored for a short period. Simulations may only need to have the source code, initial conditions, and verification data stored. In addition to explaining how data will be selected for short-term storage and long-term preservation, remember to also highlight your plans for the accompanying metadata and related code and algorithms that will allow others to interpret and use the data (see Rule 4).

Develop a sound plan for storing and protecting data over the life of the project. A good approach is to store at least three copies in at least two geographically distributed locations (e.g., original location such as a desktop computer, an external hard drive, and one or more remote sites) and to adopt a regular schedule for duplicating the data (i.e., backup). Remote locations may include an offsite collaborator's laboratory, an institutional repository (e.g., your departmental, university, or organization's repository if located in a different building), or a commercial service, such as those offered by Amazon, Dropbox, Google, and Microsoft. The backup schedule should also include testing to ensure that stored data files can be retrieved.

Your best bet for being able to access the data 20 years beyond the life of the project will likely require a more robust solution (i.e., question 3 above). Seek advice from colleagues and librarians to identify an appropriate data repository for your research domain. Many disciplines maintain specific repositories such as GenBank for nucleotide sequence data and the Protein Data Bank for protein sequences. Likewise, many universities and organizations also host institutional repositories, and there are numerous general science data repositories such as

Dryad (http://datadryad.org/), figshare (http://figshare.com/), and Zenodo (http://zenodo.org/). Alternatively, one can easily search for discipline-specific and general-use repositories via online catalogs such as http://www.re3data.org/ (i.e., REgistry of REsearch data REpositories) and http://www.biosharing.org (i.e., BioSharing). It is often considered good practice to deposit code in a host repository like GitHub that specializes in source code management as well as some types of data like large files and tabular data (see https://github.com/). Make note of any repository-specific policies (e.g., data privacy and security, requirements to submit associated code) and costs for data submission, curation, and backup that should be included in the DMP and the proposal budget.

## Rule 7: Define the Project's Data Policies

Despite what may be a natural proclivity to avoid policy and legal matters, researchers cannot afford to do so when it comes to data. Research sponsors, institutions that host research, and scientists all have a role in and obligation for promoting responsible and ethical behavior. Consequently, many research sponsors require that DMPs include explicit policy statements about how data will be managed and shared. Such policies include:

- licensing or sharing arrangements that pertain to the use of preexisting materials;
- plans for retaining, licensing, sharing, and embargoing (i.e., limiting use by others for a period of time) data, code, and other materials; and
- legal and ethical restrictions on access and use of human subject and other sensitive data.

Unfortunately, policies and laws often appear or are, in fact, confusing or contradictory. Furthermore, policies that apply within a single organization or in a given country may not apply elsewhere. When in doubt, consult your institution's office of sponsored research, the relevant Institutional Review Board, or the program officer(s) assigned to the program to which you are applying for support.

Despite these caveats, it is usually possible to develop a sound policy by following a few simple steps. First, if preexisting materials, such as data and code, are being used, identify and include a description of the relevant licensing and sharing arrangements in your DMP. Explain how third party software or libraries are used in the creation and release of new software. Note that proprietary and intellectual property rights (IPR) laws and export control regulations may limit the extent to which code and software can be shared.

Second, explain how and when the data and other research products will be made available. Be sure to explain any embargo periods or delays such as publication or patent reasons. A common practice is to make data broadly available at the time of publication, or in the case of graduate students, at the time the graduate degree is awarded. Whenever possible, apply standard rights waivers or licenses, such as those established by Open Data Commons (ODC) and Creative Commons (CC), that guide subsequent use of data and other intellectual products (see http://creativecommons.org/ and http://opendatacommons.org/licenses/pddl/summary/). The CC0 license and the ODC Public Domain Dedication and License, for example, promote unrestricted sharing and data use. Nonstandard licenses and waivers can be a significant barrier to reuse.

Third, explain how human subject and other sensitive data will be treated (e.g., see http://privacyruleandresearch.nih.gov/ for information pertaining to human health research regulations set forth in the US Health Insurance Portability and Accountability Act). Many research sponsors require that investigators engaged in human subject research approaches seek or receive prior approval from the appropriate Institutional Review Board before a grant proposal is submitted and, certainly, receive approval before the actual research is undertaken. Approvals may require that informed consent be granted, that data are anonymized, or that use is restricted in some fashion.

## Rule 8: Describe How the Data Will Be Disseminated

The best-laid preservation plans and data sharing policies do not necessarily mean that a project's data will see the light of day. Reviewers and research sponsors will be reassured that this will not be the case if you have spelled out how and when the data products will be disseminated to others, especially people outside your research group. There are passive and active ways to disseminate data. Passive approaches include posting data on a project or personal website or mailing or emailing data upon request, although the latter can be problematic when dealing with large data and bandwidth constraints. More active, robust, and preferred approaches include: (1) publishing the data in an open repository or archive (see Rule 6); (2) submitting the data (or subsets thereof) as appendices or supplements to journal articles, such as is commonly done with the PLOS family of journals; and (3) publishing the data, metadata, and relevant code as a "data paper" [5]. Data papers can be published in various journals, including *Scientific Data* (from Nature Publishing Group), the *GeoScience Data Journal* (a Wiley publication on behalf of the Royal Meteorological Society), and *GigaScience* (a joint BioMed Central and Springer publication that supports big data from many biology and life science disciplines).

A good dissemination plan includes a few concise statements. State when, how, and what data products will be made available. Generally, making data available to the greatest extent and with the fewest possible restrictions at the time of publication or project completion is encouraged. The more proactive approaches described above are greatly preferred over mailing or emailing data and will likely save significant time and money in the long run, as the data curation and sharing will be supported by the appropriate journals and repositories or archives. Furthermore, many journals and repositories provide guidelines and mechanisms for how others can appropriately cite your data, including digital object identifiers, and recommended citation formats; this helps ensure that you receive credit for the data products you create. Keep in mind that the data will be more usable and interpretable by you and others if the data are disseminated using standard, nonproprietary approaches and if the data are accompanied by metadata and associated code that is used for data processing.

## Rule 9: Assign Roles and Responsibilities

A comprehensive DMP clearly articulates the roles and responsibilities of every named individual and organization associated with the project. Roles may include data collection, data entry, QA/QC, metadata creation and management, backup, data preparation and submission to an archive, and systems administration. Consider time allocations and levels of expertise needed by staff. For small to medium size projects, a single student or postdoctoral associate who is collecting and processing the data may easily assume most or all of the data management tasks. In contrast, large, multi-investigator projects may benefit from having a dedicated staff person (s) assigned to data management.

Treat your DMP as a living document and revisit it frequently (e.g., quarterly basis). Assign a project team member to revise the plan, reflecting any new changes in protocols and policies. It is good practice to track any changes in a revision history that lists the dates that any changes were made to the plan along with the details about those changes, including who made them.

Reviewers and sponsors may be especially interested in knowing how adherence to the data management plan will be assessed and demonstrated, as well as how, and by whom, data will be managed and made available after the project concludes. With respect to the latter, it is often sufficient to include a pointer to the policies and procedures that are followed by the repository where you plan to deposit your data. Be sure to note any contributions by nonproject staff, such as any repository, systems administration, backup, training, or high-performance computing support provided by your institution.

## Rule 10: Prepare a Realistic Budget

Creating, managing, publishing, and sharing high-quality data is as much a part of the 21st century research enterprise as is publishing the results. Data management is not new—rather, it is something that all researchers already do. Nonetheless, a common mistake in developing a DMP is forgetting to budget for the activities. Data management takes time and costs money in terms of software, hardware, and personnel. Review your plan and make sure that there are lines in the budget to support the people that manage the data (see Rule 9) as well as pay for the requisite hardware, software, and services. Check with the preferred data repository (see Rule 6) so that requisite fees and services are budgeted appropriately. As space allows, facilitate reviewers by pointing to specific lines or sections in the budget and budget justification pages. Experienced reviewers will be on the lookout for unfunded components, but they will also recognize that greater or lesser investments in data management depend upon the nature of the research and the types of data.

## Conclusion

A data management plan should provide you and others with an easy-to-follow road map that will guide and explain how data are treated throughout the life of the project and after the project is completed. The ten simple rules presented here are designed to aid you in writing a good plan that is logical and comprehensive, that will pass muster with reviewers and research sponsors, and that you can put into practice should your project be funded. A DMP provides a vehicle for conveying information to and setting expectations for your project team during both the proposal and project planning stages, as well as during project team meetings later, when the project is underway. That said, no plan is perfect. Plans do become better through use. The best plans are "living documents" that are periodically reviewed and revised as necessary according to needs and any changes in protocols (e.g., metadata, QA/QC, storage), policy, technology, and staff, as well as reused, in that the most successful parts of the plan are incorporated into subsequent projects. A public, machine-readable, and openly licensed DMP is much more likely to be incorporated into future projects and to have higher impact; such increased transparency in the research funding process (e.g., publication of proposals and DMPs) can assist researchers and sponsors in discovering data and potential collaborators, educating about data management, and monitoring policy compliance [6].

## Acknowledgments

This article is the outcome of a series of training workshops provided for new faculty, postdoctoral associates, and graduate students.

## References

1. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comput Biol 2(2): e12. PMID: 16501664
2. Zhang W. (2014) Ten simple rules for writing research papers. PLOS Comput Biol 10(1):e1003453. doi: 10.1371/journal.pcbi.1003453 PMID: 24499936

3. Bourne PE. (2007) Ten simple rules for making good oral presentations. PLOS Comput Biol 3(4): e77. PMID: 17500596
4. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, et al. (2014) Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol 10(4): e1003542. doi: 10.1371/journal.pcbi.1003542 PMID: 24763340
5. Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinf 12 (Suppl 15):S2.
6. Mietchen D (2014) The transformative nature of transparency in research funding. PLoS Biol 12(12): e1002027. doi: 10.1371/journal.pbio.1002027 PMID: 25549343

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Experiments' Provenance

**Toni Kazic***

Dept. of Computer Science Missouri Maize Center, Missouri Informatics Institute, and Interdisciplinary Plant Group, University of Missouri, Columbia, Missouri, United States of America

* kazict@missouri.edu

Everyone needs experimental data to understand biology. Exactly how and from what the data were obtained determines an experiment's results, specifies how it can be reproduced, and conditions our analyses and interpretations. These details of materials, methods, and analyses are the experiment's provenance.

Today, as it has been for hundreds of years, experimental provenance is recorded in some form of laboratory notebook. But as data migrate from the experimentalist's mind and notebook to publication, the lab server, the archival database, or the cloud, this essential information now vanishes. Like interpretation, our ability to reproduce results depends on knowing how they were produced by others. Shorn of their immediate context, the methodological ideas and information that were perfectly transparent to the experimentalist (or computationalist!) become opportunities for error-prone reconstruction by others, even within the same group [1–3]. That reconstruction requires (repeated) private communications, rereading notebook entries, polling one's own or a group's collective memory, and looking at the specimens. None of those methods are reliable, and all are tedious.

As big data become a reality, it will be ever more imperative to encapsulate experimental provenance with the data. But how do we get that information out of the brains and notebooks in the first place? This is a problem of information capture, not data formats; of laboratory practice, not resource discovery; and of the million flowers of experimental creativity, not ontology building. Of course, ontologies, interoperable grid resources, and efficient search are important and appealing—but absent experimental provenance, they are biologically moot.

The obvious solution might seem to be standards for post hoc data annotation by biologists. Indeed, several previous efforts have defined sets of "minimal" metadata about particular types of high-throughput experiments, beginning with the minimal information needed for microarray experiments (the MIAME criteria) [4]. However, experience shows three fundamental problems with this approach. First, despite vigorous encouragement from computational biologists, most deposited datasets lack such annotation [1]. Second, the universe of experiments performed, let alone possible, far exceeds the stamina of even the most earnest committees to promulgate definitions and criteria. As always, science outruns nomenclature.

The third fundamental problem is even more basic than effort or invention. Many experimentally inclined biologists are too ill equipped and too busy to produce electronic provenance metadata in almost any form. Large experimental consortia and high-throughput facilities often do develop in-house provenance systems, and these and commercial ones are available (for example, see [5,6]). But smaller groups also generate essential data, often with a wide variety of experiments that don't fit existing packages, standards, or ontologies. Their provenance information tends to be fragmented, buried in a mix of paper and electronic records, and dependent on the



**Citation:** Kazic T (2015) Ten Simple Rules for Experiments' Provenance. PLoS Comput Biol 11(10): e1004384. doi:10.1371/journal.pcbi.1004384

**Published:** October 20, 2015



**Funding:** I gratefully acknowledge the support of the U.S. NSF (www.nsf.gov), MCB-1122130 and the University of Missouri Research Board. The funders had no role in the preparation of this manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

group's institutional memory. Provenance is especially important in these contexts, since many of these experiments will nucleate the hypotheses and provide the materials for the subsequent high-throughput experiments that are the meat and potatoes of much of contemporary computational biology. While metadata are notoriously difficult to obtain from experimentalists, in principle they could be computed from adequate electronic provenance records. The first difficulty in that sentence is "adequate electronic provenance records." Like other forms of housekeeping, it is easy for provenance to be mere drudgery, without the glass slipper at the end.

How can we make provenance easier so it is better for all? The trick is to capture provenance as the experiment is planned, performed, and analyzed. The easier, more familiar, and more helpful to the experimentalist capture is, the more routine it can become. Now, "easy" is the toughest design goal of all, and building general systems is hard and expensive. But what could we do in an afternoon or a few? I think the answer is quite a lot, provided the "we" is a joint effort of the experimentally and computationally inclined. Experimentalists must repeatedly show what they do, explain how they think, and critically test prototypes. Computationalists must repeatedly observe all the acts of experimentation, listen for unstated assumptions, and prototype the least intrusive, most experimentally efficient approaches. Together, both should maximize simplicity, flexibility, extensibility, and fun.

Capturing the experimental record at the source in real time in all types of laboratories will smooth the path to systems that automate capture and combine it with the extraction of provenance and annotatation of datasets. To reach the provenance infrastructure of the future that everyone needs, we need to understand the diversity of actual experimental practice and to start solving that most difficult problem of provenance capture. A brief dollop of altruism that focuses on very quick, lightweight, shareable improvements could immediately help experimentalists, provoke engagement across boundaries, and seed more sustainable collaborations. So, in the spirit of the rules for the provenance of computational experiments and instrument data [7,8], I offer ten simple rules for interdisciplinary collaborations on provenance capture. The order of the rules roughly parallels the workflow of discussions, from that first exploratory cup of coffee to the migration to grander schemes and bigger data. While each rule varies a bit in the distribution of tasks to experimentalists and computationalists, all require a joint effort. Our current practice is described in Box 1.

## Rule 1. Go Sideways and Backward to Go Forward

What do the experimentalists track now, and what physical items and ideas interact with that? Similarly, many materials and methods have a history that is crucial to capture: genetic pedigrees and macromolecule preps are examples. So, discuss each major phase of the experimental lab's life and how those phases relate to its work of today and tomorrow. Retrospectively entered data then join a consistent framework, rather than being kludged in.

## Rule 2. Improve the Acts of Experimentation

No one willingly adds encumberances, so changes have to produce a net gain over the entire experimental workflow. People are often willing to sacrifice a few person-days to save person-months, but to identify improvements is a joint labor of reengineering. Multiple interactions about and observations of the "same" experimental task show the ways in which the work varies, pinpointing process improvements and delimiting a design's flexibility.

## Rule 3. It's Gotta Beat a Spreadsheet

Spreadsheets are ubiquitous because they're flexible, well understood by a large community (including students), great for prototyping an experimental workflow, and collect data simply.

## Box 1. Our Practice So Far

As a computational biologist who has also done maize genetics for the past nine years, I have the privilege of directly experiencing the realities of my experimentalist colleagues while watching how well—or not!—my computational ideas address the practical problems of experiments' provenance. We study a set of ≈55 distinct maize mutants that produce necrotic or chlorotic lesions on leaf tissue [13,14]. We package selected seed for planting; plant in a field at research farms or in pots or trays in the greenhouse; repeatedly observe at least eight different phenotypes for each plant; pollinate with selected plants; photograph leaves in situ or ex situ; orally describe each mutant family; collect, lyophilize, and freeze samples of leaf tissue for DNA sequencing; harvest and shell pollinated ears; and file the corn for easy retrieval from the cold room. Field data are collected in the form of images (either standardized or free-form), spreadsheet tables, audio recordings, dumps of hourly weather data from a local recording station, and geographic coordinates of the first row of each field.

How do we apply the rules? Our simple provenance system for maize genetics has gradually evolved. The basic system was designed in the course of the first crop, with much discussion with maize colleagues about what and how they track their provenance. The result has proven very versatile and robust, needing only minimal changes despite the changing circumstances of each crop and the addition of other experiments and data and object types. Robustness is difficult to quantitate, but the system has so far managed approximately 5,900 families of maize (4,000 in active use), 18,000 images (including duplicates and test shots), 1,000 tissue samples, and 430,000 facts (including both primary data and reverse indices). We have added experiments and experimental protocols over the years, changed key equipment more than once, and worked with approximately 20 students so far in the project. These volumes are small compared to those of many experimental groups, especially in maize, but may suggest our experimental milieu.

The heart of our system is the unique identifiers. *Every* physical object that contributes directly to the production of our biological materials or for which data are collected has a unique identifier. There are many types of objects involved, and we often need to know what to do quickly, so we use mnemomic identifiers that distinguish each type of object and distinguish plants and their progency from equipment. Standard equipment for the field and seed room are not tracked—staplers, aprons, and shellers are all interchangeable. Cameras and scanners are not, so each has a distinctive name. So far, lenses are permanently assigned to cameras, and camera names are recorded as part of the photographic data. If we were to exchange lenses among cameras, it would be simple to name the lenses and their associations with cameras in the provenance system so that past and present data were correctly annotated.

Each plant is tagged with a sturdy barcoded paper strip that has multiple tear-off tags, each printed with the plant's identifier, bar code, and an abbreviated symbolic genotype. The plant's identifier becomes the primary key for all tabular data and seed from that plant, and it is the plant's name in audio narratives, linking genotypes, phenotypes, samples, and data. Plant identifiers are 15 characters and state the year of planting, season, family number, inbred background if relevant, and the row and plant number for each plant. Redundancy is built into the identifier to help guard against information loss. Pollinations are labelled with tear-off tags from the plants serving as female and male for that particular cross. Stapled together and to the seed envelope, they identify the shelled seed for inventory, retrieval, and packing. The few person-days spent tagging the thousands of plants in each crop saves many person-months in data collection and verification, inventory management, and computation.

Pots, trays, seed packets, row stakes, tissue samples, boxes, sleeves, and seed bags all receive unique six-character identifying bar codes, with the identifiers for each type of object beginning with a single mnemomic letter. All letters, including those for inbred lines, are unique. Leaves are identified by either a relative or absolute coordinate on the plant, depending on the experiment. Apart from the leaf identifiers, all identifiers and their components are automatically generated, a lesson learned in the second year of field work when identifiers for a few families of siblings were inadvertantly duplicated.

Identifiers are printed in a large, bold font, along with their one-dimensional bar codes, on labels or tags. Labels and tags are generated with custom scripts and open-source code [9,10,15,16]. Our script collection includes code to generate individual tags to replace those with worn, illegible bar codes, hastily repurposed tags from sibling plants, and tags with retrospectively corrected data.

*Every* action or datum involving a barcoded object is recorded by scanning the bar code into a data table in a spreadsheet, either at the moment of the action or shortly thereafter. Contemporaneous data collection is one of our best safeguards against mangled data, permitting correction while the object or action is immediately present. It also helps us spot procedural bottlenecks and error-prone operations for process improvement. The only exception to the rule of contemporaneous data collection is for intermediate forms of the data, such as emerging lesion phenotypes or the pollination and photographic plan for a plant. These are stored on each plant as color-coded paper twist ties, with the date and initials of the human scorer stored on the first plant of the row. (We do record dates and scorer of each plant on it as needed, for example, when determining the onset of phenotypes.) These decisions can change as the phenotypes develop and pollinations proceed, so we usually record only the final evaluation or intention.

Most data are collected by scanning bar codes into a spreadsheet running on a tablet, using a matchbox-sized bluetooth scanner. Representative leaves from selected mutant plants are photographed to record phenotypes and to provide data for their quantitative characterization. Other photographs compare phenotypes among families and document surprises. Audio recordings of descriptions of the field, crop, families, and individual plants are collected throughout the field season, formerly with various dictaphone arrangements and now with the tablets. Their transcription lags, so we are now experimenting with speech-to-text programs.

Conversion of data from spreadsheet to database uses a family of Perl scripts and modules, including a library of regular expressions. As our data collection machinery has migrated from menu scanners to iPads and spreadsheets, and as students have come and gone, the characteristic errors that appear during data collection have changed. We dump the data as csv files and manually check those files before processing them and inserting their data into the database. Each student reviews the data he or she collected, and we also review each other's data. We also perform different post hoc checks, depending on the operation—making sure each row and packet are accounted for at planting, that all recorded and unrecorded pollinations are harvested, that every ear used in pollinations is unique, etc.

Computationally, our provenance system uses a mix of tools: a declarative database for crop and data management, including pedigree computations; emacs org-mode, for the lab's notebooks; git and tar for archiving ASCII data and code; and Perl scripts for generating tags and labels, generating org-mode tables with embedded, readily visible calculations, and for converting data from spreadsheet dumps and org files to the database. All types of files are backed up on two physically different RAID arrays. The experimental provenance system was developed before our system for computations and analyses, so the two interact at several levels without forming a monolith.

Images, files, and file and directory names all self-identify. Our leaf images include a barcoded tag from the plant, marked with the leaf number. (This practice has rescued data from scanning errors more than once.) All ASCII files begin with a string that includes the file's full path. Files or data increments produced by code include comments specifying the name of the producing file or function, the source file for the data, and the timestamp of production. File names for audio recordings are descriptive now that we collect these with a tablet instead of a dictaphone. Names in directory trees are descriptive; camera and scanner names form part of the directory tree for primary data storage, helping us rapidly locate the data referenced in tables. Our lab notebook was formerly a set of physical notebooks and ASCII files. We recently switched to emacs org-mode, which facilitates project management and publication as well as written narratives. We photograph whiteboards, oddities, and paper, cross-referencing the images in our org-mode files and filing them in the same directories as the work they reference. We occasionally record conversations, and these are cross-referenced and filed in the same way.

Data and computation semantics reside in predicate and argument names and comments in files, but the more complex semantics still live in text files or my brain. Our next provenance frontier is to compute our metadata more easily, starting with experimental images intended for public deposition. Current ontologies denote only a tiny fraction of what we deem important, but this may change in the future.

That combination is hard to beat! Yet, minor innovations in spreadsheets can yield big improvements in provenance. Examples include restructuring repeated free text descriptions as menu items and providing optional pop-up boxes with definitions of the lab's terms and methods. More extensive systems with designer interfaces or back ends can be merited if the experimental workflows are very regular, but any proffered replacement should be as simple and as robust as a spreadsheet to use and maintain. Homebrew systems without a trivial maintenance path die once the graduate student who built them moves on.

## Rule 4. Barcode Everything Important and Keep the Labels Current

Each type of physical object or datum should have a distinctive, mnemomic identifier that tells you what it is without needing a reference guide or a gadget. What should be memorialized and how much information should be incorporated into the identifier for optimal tracking depends on the laboratory, and designing good identifier systems that are robust to change takes care. Mnemomic identifiers are easier to use in everyday experimental discourse than unadorned integers but may need more maintenance, especially in the face of the inevitable revisions. When the laboratory has many different types of physical objects and the context of their relationships is important in knowing what to do (usually quite urgently), then mnemomic identifiers can be a great help. They also have the advantage of letting one embed redundant information into the identifier. When the objects form relatively few types and the relationships among them are as yet unknown, then a centrally assigned integer can prove simpler in the long run. In either case, it is crucial to avoid embedding any (often subconscious) biological assumptions or inferences in the identifier. Similarly, storage systems and the organization of collections change over time. Rather than build fungible relationships and inferences into the identifier, identify the bone and the drawer it's in today separately; store the encoded site of the bone's collection in a database, revising the site's coordinates when Global Positioning System (GPS) data are substituted for sunsights; and discover the relationships among bones collected at the same site by experiments.

Hastily improvised or newly inadequate identifiers are a fact of life and may not be transformed into the standard scheme for some time. If any part of the identifier for the object changes in the database (that site has more bones than we have characters in the identifier!), print the new identifier and its bar code on a label and attach it to the object so that both new and old labels can be read as needed. Keeping the labels current with the inevitable changes in the databases prevents the confusion that results from scanning old labels into new data schemes, minimizing repairs.

Once the identifier scheme is worked out, producing good labels is easy. Both open-source and commercial programs to generate bar codes are available, and it's easy to write one's own [9,10]. Print the identifier in large, bold font next to the bar code on sturdy label or tag stock so that a glance tells the story. There are a variety of materials, tags, and labels that are waterproof, take ink well without smearing, and tolerate temperature and humidity extremes so that plates and tubes can be barcoded.

## Rule 5. Make Everything Self-Identifying

Any bar code can be mangled on scanning or land in the wrong place in the spreadsheet. Looking at the physical artifact or data file can resolve these problems, but only if those things self-identify. Self-identification can be as simple as descriptive directory and file names, including file names as the first line in text files, or photographing a labelled container or rack of tubes. Though namespace collisions can occur with descriptive names, they are far more legible than random strings. A contemporaneous record guards against all errors except mislabelling the original.

## Rule 6. Use Version Control and Backups

Electronic laboratory notebooks are available in both open-source and commercial versions, and some may fit a group's workflow well enough to justify the effort and expense of adoption. However, for many, a mnemomically named directory scheme, spreadsheets, text files, version control, and a RAID array may be enough—and a significant improvement. Version control of directories with ASCII data and notes is a cheap way to archive and time-stamp changes, emulating electronic notebooks while preserving workflow flexibility. It also provides the most insurance against fat fingers, fatigue, and forgetfulness. Calling export and backup scripts from a cron job or a big green "archive now!" button reduces human supervision.

## Rule 7. Automate Gluing over Cleaning

It's usually not worth investing in sophisticated error correction much beyond essential reformatting. Many errors are idiosyncratic and disappear with practice in cleaning one's data. Cleaning is painful, even with macros, but it teaches one to minimize collection errors, increases attention to experimental details, and creates another opportunity to check for lurking substantive errors. Rather, automate data transformations and archiving, and review characteristic errors from time to time. A similar principle applies to data generated in high-throughput facilities. If its managers are amenable, automating data transfers from a facility to the ultimate storage device saves time and error.

## Box 2. A Glimpse of the Landscape

An abyss separates the practices in many laboratories and research on the semantics and provenance of data and computations, workflow systems, and electronic notebooks and groupware. Beyond the simple rules, scalably bridging the divide will require connecting today's achievements into flexible, transparent, and interoperable ecosystems of applications that meet *experimentalists'* felt needs. Of course, we have seen this landscape before: a visionary system that addressed many of these issues was the Worm Community System of the early 1990s [17,18]. Perhaps the following sketchy list can stimulate some collaborative spanning.

### Metadata and provenance

In the rules, I emphasized provenance acquisition, skirted the formalization of metadata, and ignored provenance maintenance and reconciliation [19–26]. Archivists call information about the origin and semantics of experimental objects, computations, datasets, and analyses "descriptive metadata" (a book's content), distinguishing them from the "administrative metadata" needed to use and manage resources (a book's library record). However, many desirable applications would use both notions transparently. The early successes of the Dublin Core (administrative metadata), Gene Ontology, and macromolecular crystallographic information file (mmCIF) (the latter two, descriptive metadata) encouraged the idea that metadata would naturally arise from ontologies and be exchanged through the web [27–29]. The result was a flowering of standards, societies, ontologies, and working groups, each aimed at a particular slice of biology. Much of this work is now represented in database annotations, ontology and Semantic Web languages, and projects that foster and archive these materials [30–32]. Nonetheless, experimentalists treat annotations as read-only data and are largely unaware of the rest of the infrastructure. The *Nature* methods checklist is an interesting mix of metadata and location information [33,34]. It may further stimulate work on automated extraction of scientific, descriptive metadata—and validation of the output!—as is now done for bibliographic, administrative information [35–40]. We might even hope for a day when datasets and computations have unique identifiers, rather like DOIs, to facilitate building chains of provenance in both senses.

### Workflow systems

Increasingly, wet-bench and computational work form an integrated whole, but current workflow systems address either a portion of the wet-bench work—such as LabView's abilities to interconnect multiple instruments and their data—or computations [41]. At the moment, connections between the two sides mostly reside in the experimenter's brain or in his or her notebooks. Research so far has been mainly on the essential technical details of organizing, tracking, and managing data, code, cycles, and storage [42–47]. Heroic efforts are made to stimulate adoption by each system's notional user communities, but the need for heroism suggests that we should watch many experimentalists more carefully to uncover their desires and reservations.

### Electronic notebooks and groupware

Cross literate programming, record keeping, project management, and text editors. Their offspring range from the free and open-source ipython notebook and emacs org-mode aimed at individual investigators, to commercial systems for groups, to cloud-based editors such as Evernote and SimpleNote, to wrangling group documents in wikis, GoogleDocs, or SharePoint [48–51]. They vary enormously in their support of all the desiderata mentioned, as do the needs of research groups for these good things. It may be time to consider how open architectures could form modular, niche environments that also ease the experimentalists' journeys to thoughtful computation.

## Rule 8. Integrate the Paper

Many brains, even young ones, think more fluently in front of paper (or whiteboards) than screens. Photograph or scan these, cross-referencing the images in whatever the group uses for electronic notebooks and filing them in the appropriate directories. Groups that use just a few forms of stationary may benefit from switching to paper random-dot or grid notebooks and recording pens, but this may be too restrictive or expensive for others. Don't forget the other form of "paper" that is produced by thinking out loud, and capture that with audio recordings. Common smartphone and tablet operating systems all run free voice recorder apps. Some groups may be able to use speech-to-text systems successfully to collect data.

## Rule 9. Prepare to Extract Metadata

Spreadsheet field headers, database attributes, and free text notes are crude metadata. Scooping them into a database that indexes their location will facilitate eventual extraction and lets one track other information more easily than recursive greps. The scoops need be no fancier than a text box on a form, a script that parses a csv or Excel file, or names of pdf files of paper notebook pages describing an experiment's method [11]. More sophisticated tools that embed metadata collection in spreadsheet templates are also available [12]. Watch for repeated phrases, which are good metadata candidates.

## Rule 10. Only the Biologists Know for Sure

The harder part of metadata is ensuring the data's actual and declared semantics match. The experimentalist is the only one who knows, and intermittent discussion will reveal crucial subconscious information. It follows that the definitions of metadata terms must be available at the moment of entry. A box for free text insertion lets one capture and analyze emerging needs, given a more committed collaboration.

Should you use existing standards, ontologies, and metadata (Box 2)? Of course, if these conventions are stable, capture what the experimentalist needs to say, and the experimentalist agrees with the convention's semantics. Ontologies in rapid-flux, fuzzily defined terms or odd lumpings of tradtional nomenclature are not good candidates for describing experiments. Then it is particularly important to record how the experimentalist describes the experiment and its data and to transmit that information to ontologists so they have more usages to study, while together you use them to define new metadata. As complex data accumulate in public resources, those resources will have to manage the migration of ontological terms. For now, free text seems the best guide to accurate migration, albeit the slowest.

## Acknowledgments

I'm deeply grateful to my students, who have field- and lab-tested these ideas and forced their steady improvement, often despite adversity. None of this work would have been possible without my maize colleagues, at Missouri and around the world, who patiently taught me the ropes of maize genetics and their management systems. Sanda Erdelez, Dietlind Gerloff, Derek Kelly, Janet Kelso, Mac, Susan Melia-Hancock, Heather Moulaison, Brian Pitts, Mary Schaeffer, Ann Stapleton, Vinny, and William Wise made many helpful suggestions and are completely blameless.

## References

1. Begley CG, Ellis LM (2012) Raise standards for preclinical cancer research. Nature 483: 531–533. doi: 10.1038/483531a PMID: 22460880
2. Blow NS (2014) A simple question of reproducibility. Biotechniques 56: 8. PMID: 24592483
3. Garijo D, Kinnings S, Xie L, Xie L, Zhang Y, et al. (2013) Quantifying reproducibility in computational biology: the case of the tuberculosis drugome. PLoS One 8: e80278. doi: 10.1371/journal.pone.0080278 PMID: 24312207
4. FGED (2015—present). MIAME. Minimum information about a microarray experiment. Functional Genomics Data Society. http://fged.org/projects/miame/
5. Sanchez-Villeda H, Schroeder S, Polacco M, McMullen M, Havermann S, et al. (2003) Development of an Integrated Laboratory Information Management System for the Maize Mapping Project. Bioinformatics 19: 2022–2030. PMID: 14594706
6. Hernández de Diego R, Boix-Chova N, Gómez-Cabrero D, Tegner J, Abugessaisa I, et al. (2014) STATegra EMS: an Experiment Management System for complex next-generation omics experiments. BMC Sys Biol 8(Suppl 2): 59.
7. Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten simple rules for reproducible computational research. PLoS Computational Biology 9: e1003285. doi: 10.1371/journal.pcbi.1003285 PMID: 24204232
8. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, et al. (2014) 10 simple rules for the care and feeding of scientific data. PLoS Computational Biology 10: e1003542. doi: 10.1371/journal.pcbi.1003542 PMID: 24763340
9. Rubini A (1999–present) GNU Barcode. http://www.gnu.org/software/barcode/. Free Software Foundation.
10. Kazic T (2006–present) Welcome to maizelesions.org! http://www.maizelesions.org/. University of Missouri, Columbia.
11. Wilson, D (2015—present) Spreadsheet::ParseExcel. http://search.cpan.org/~dougw/Spreadsheet-ParseExcel-0.65/lib/Spreadsheet/ParseExcel.pm
12. RightField Developers (2008–present) RightField. http://www.rightfield.org.uk.
13. Neuffer MG, Calvert OH (1975) Dominant disease lesion mimics in maize. J Heredity 66: 265–270.
14. Neuffer MG, Coe Edward H Jr, Wessler SR (1997) Mutants of Maize. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
15. Lamport L (1994) LaTeX: a Document Preparation System. Reading MA: Addison-Wesley Publishing Co., second edition.
16. Artifex Software, Inc (1989–present) Ghostscript. www.ghostscript.com: ghostscript.com.
17. Schatz BR (1992) Building an electronic community system. J Managemt Info Sys 8: 87–107.
18. Pool R (1993) Beyond databases and email. Science 261: 841–843. PMID: 8346436
19. Simmhan YL, Plale B, Gannon D (2005) A survey of data provenance in e-science. SIGMOD Rec 34: 31–36.
20. Buneman P, Davidson SB (2010) Data provenance—the foundation of data quality. Technical report, University of Pennsylvania, Philadelphia.
21. Bowers S, McPhillips T, Ludäscher B, Cohern S, Davidson SB (2006) A model for user-oriented data provenance in pipelined scientific workflows. In: International Provenance and Annotation Workshop (IPAW). Berlin: Springer Verlag. *Lec*. *Notes Comp*. *Sci*. 4145: 133–147.
22. Bourne PE, Clark T, Dale R, de Waard A, Herman I, et al. (2012–present) Improving future research communication and e-scholarship. http://www.force11.org/white_paper.
23. Force11 (2014–present) Force11. https://www.force11.org.
24. Lebo T, Sahoo S, McGuinness D, Belhajjame K, Cheney J, et al. (2013–present) PROV-O: The PROV ontology. W3C Recommendation 30 April 2013. http://www.w3.org/TR/2013/REC-prov-o-20130430/. W3C.

25. Data to Insight Center (2014–present) Data to Insight Center. http://d2i.indiana.edu. University of Indiana.

26. Hedstrom M, Alter G, Kumar P, Inna K, McDonald RH, et al. (2013) SEAD: An integrated infrastructure to support data stewardship in sustainability science. In: CASC Research Data Management Implementation Symposium, March 13–14, 2013, Arlington, VA. Arlington, VA: CASC, p. http://dx.doi.org/10.6084/m9.figshare.651719.

27. Dublin Core Metadata Initiative (2014–present) Dublin Core Metadata Initiative. Making It Easier to Find Information. http://dublincore.org.

28. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. Nature Genet 25: 25–29. PMID: 10802651

29. wwPDB, PDBx/mmCIF Dictionary Resources, Research Collaboratory for Structural Biology, 2015—present. http://mmcif.wwpdb.org/

30. Patel-Schneider PF, Hayes P, Horrocks I (2004) OWL Web Ontology Language Semantics and Abstract Syntax Section 5. RDF-Compatible Models. http://www.w3.org/TR/owl-semantics/rdfs.html. W3C.

31. National Center for Biomedical Ontology (2005) OBO: Open Biomedical Ontologies. http://obo.sourceforge.net/.

32. Goble CA, Stevens R, Ng G, Bechhofer S, Paton NW, et al. (2001) Transparent access to multiple bioinformatics information sources. IBM Syst J 40: 532–552.

33. Nature Publishing Group (2013) Reporting checklist for life sciences articles. Nature 496: 398.

34. Nature Publishing Group (2013) Reporting checklist for life sciences articles. http://www.nature.com/authors/policies/checklist.pdf.

35. Lin S, Ng J, Pradhan S, Shah J, Pietrobon R, et al. (2010) Extracting formulaic and free text clinical research articles metadata using conditional random fields. In: Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, Los Angeles, June 2010. Los Angeles: Association for Computational Linguistics, pp. 90–95.

36. Hespanha SR (2013–present) Text Mining for Ontologies (TMO). http://www.nceas.ucsb.edu/hespanha/srh/Projects/Entries/2013/2/22_text_mining_for_ontologies_%28TMO%29.html. NCEAS.

37. Rzhetsky A, Evans JA (2011) War of ontology worlds: mathematics, computer code, or Esperanto? PLoS Computational Biology 7: e1002191. doi: 10.1371/journal.pcbi.1002191 PMID: 21980276

38. Bandrowski A, Cachat J, Li Y, Muller H, Sternberg P, et al. (2012) A hybrid human and machine resource curation pipeline for the Neuroscience Information Framework. Database 2012: bas005. doi: 10.1093/database/bas005 PMID: 22434839

39. Funk C, Baumgartner W Jr, Garcia B, Roeder C, Bada M, et al. (2014) Large-scale biomedical concept recognition: an evaluation of current automatic annotatorsand their parameters. BMC Bioinfo 15: 59.

40. Dutkowski J, Kramer M, Surma MA, Balakrishnan R, Cherry JM, et al. (2013) A gene ontology inferred from molecular networks. Nature Biotechnol 31: 38–45.

41. National Instruments (2014–present) LabView System Design Software. http://www.ni.com/labview/.

42. iPlant Collaborative (2008–present) iPlant Collaborative. Empowering a New Plant Biology. http://iplantcollaborative.org/. Cold Spring Harbor Laboratory and University of Arizona.

43. Taverna Developers (2007) Taverna. http://taverna.sourceforge.net/?doc = download.html.

44. myexperiment Developers (2014–present) myexperiment. http://www.myexperiment.org/.

45. Kepler Collaboration (2007) Kepler Project. http://kepler-project.org.

46. LONI Developers (2014–present) LONI Pipeline. http://pipeline.loni.ucla.edu/.

47. DataONE Developers (2014–present) DataONE. Data Observation Network for Earth. http://www.dataone.org/.

48. Schulte E, Davison D (2011) Active documents with org-mode. Comput Sci Eng 13: 66.

49. Pérez F, Granger BE (2007) ipython: a system for interactive scientific computing. Comput Sci Eng 9: 21–29.

50. iPython Development Team (2014–present) iPython Notebook. http://ipython.org/notebook.html.

51. Giles J (2012) Going paperless: the digital lab. Nature 481: 430–431. doi: 10.1038/481430a PMID: 22281576

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Protecting Research Integrity

**David M. Shaw[1]*, Thomas C. Erren[2]**

1 Institute for Biomedical Ethics, University of Basel, Basel, Switzerland, 2 Institute and Policlinic for Occupational Medicine, Environmental Medicine and Prevention Research, University Hospital of Cologne, Cologne, Germany

* david.shaw@unibas.ch

OPEN ACCESS

**Citation:** Shaw DM, Erren TC (2015) Ten Simple Rules for Protecting Research Integrity. PLoS Comput Biol 11(10): e1004388. doi:10.1371/journal.pcbi.1004388

**Editor:** Scott Markel, Accelrys, UNITED STATES

**Published:** October 1, 2015



**Funding:** There was no special funding associated with this paper.

**Competing Interests:** The authors have declared that no competing interests exist.

Research integrity is frequently highlighted as an essential component of modern medicine and science. Adherence to the ethical principles of one's profession might seem like a simple task, but research misconduct remains a serious problem. Despite repeated calls for increased emphasis on the importance of research integrity [1–6] and a proliferation of guidelines regulating scientific misconduct at the international, national, and institutional levels [7,8], recent scandals concerning falsification and suppression of results [9,10] suggest that we need to more carefully nurture the ethical integrity of our research endeavours.

Most of the recent controversy concerning scientific misconduct has focused on plagiarism and fabrication of results. This type of malpractice has rightly been universally recognized as a very serious breach of research integrity. However, the focus on these abuses has distracted attention away from other practices which, while they may not jeopardize the scientific process to the same extent, are nonetheless clearly breaches of scientific integrity. Much more could be done to combat authorship misattribution, failure to declare relevant conflicts of interest (COIs), failure to report results, and misuse of metrics in funding decisions—in terms of both establishing stronger guidelines and ensuring their enforcement.

Although many institutions have official integrity guidelines, these frequently act as a window dressing—claiming to address key integrity issues while allowing corrosive low-level misconduct to proliferate. For example, putting pressure on junior researchers to allow guest authors on papers is not only unethical but also makes it more likely that they will see such behaviour as acceptable. Here, we suggest ten simple rules that should be put into action by all research institutions and to which all researchers should adhere in order to ensure ethical behaviour in science and medicine. In our view, all integrity guidelines should mandate these minimum requirements. Rules 1–5 are specific recommendations for particular integrity issues, and Rules 6–10 concern what institutions themselves should do.

## Rule 1: All Papers Submitted Must Contain Contributorship Statements

One of the most well-known integrity-related issues is authorship. Everyone in academia has heard of "guest" [11] and "ghost" [12] authors, and most integrity guidelines now mandate adherence to international guidelines and forbid dishonest authorship attribution [13]. This is fine in principle but insufficient in practice; institutional authorship guidelines often differ from those of journals, and even widely accepted guidelines contain contradictions (for example, the International Committee of Medical Journal Editors [ICMJE] criteria offer two different definitions of authorship [14]). In fact, the existence of authorship and research integrity guidelines may provide a false image of respectability for academia [14], as the numerous conflicts of interest that afflict research institutions make it very difficult for both junior and senior researchers to adhere to such guidelines [15]. For example, if a researcher suspects misattribution of authorship, it is his duty to raise concerns, but this poses a clear COI: if he does, he might risk his job or even his career.

What else can be done to ensure honest authorship attribution? The bar must be raised: integrity guidelines should mandate contributorship statements, which are widely recognized as a much more transparent means of attributing authorship and also have the advantage of avoiding ambiguity about which of several conflicting guidelines are being followed [16]. While it is relatively easy (and erroneously recommended in a reputable journal like *Nature* [11,17,18]) for researchers to add their department head as an author when he didn't contribute anything to the paper, or even to lie on the declaration form and say that he did in fact contribute, it is much more difficult to actively lie on a contributorship statement. The concept of using statements to make contributors accountable was proposed almost 20 years ago [12]; some journals (including the *Journal of the American Medical Association* [*JAMA*] and the *British Medical Journal* [*BMJ*]) do require contributorship statements, and the ICMJE guidelines recommend their use as well [19]. However, these statements are still not mandated by any major integrity guideline, meaning that researchers are unlikely to be sanctioned by their institutions for failing to state contributions accurately. This is a typical example of paying lip service to research integrity without actually speaking to the problem.

In order to increase transparency, all authors should put contributorship statements in every paper they submit, regardless of journal policies. Journals that refuse to print such statements should be considered blacklisted by ethical researchers. Given that ghost and guest authorship are in a similar category as plagiarism, it is surprising that they are not treated with the same gravity. Ideally, contributorship statements would replace traditional author lists at the start of articles, rather than being buried at the end of the paper, which still allows assumptions about author order to play a role. Such statements are important not only in order to acknowledge non-author contributions but also to make it clear exactly who should take the credit and who bears the responsibility for different aspects of a research paper, rather than letting readers guess based on the order of authors.

Some studies have already investigated whether contributorship statements impart relevant information [20], but we need more empirical evidence regarding how researchers are obeying the many available rules.

## Rule 2: All Financial COIs Must Be Reported with No Time Limits

As well as being honest in terms of authorship, researchers are expected to declare all conflicts of interest when applying for funding, publishing papers, or peer reviewing. Nonetheless, many COI policies and COI clauses in integrity guidelines are weak; even supposedly "comprehensive" checklists can omit items such as previous personal fees paid by sponsors [21]. Requiring a declaration of current financial COIs or only those from the last few years is insufficient when research has shown that even receiving free pens can influence physician decision-making [22]. All financial COIs that could ever have affected a researcher's judgement should be declared, with no time limits.

## Rule 3: Relevant, Non-financial Potential COIs Must Be Declared

Furthermore, more emphasis should be put on non-financial COI, which can bias researchers just as much as money [23,24]. *The Lancet* has perhaps the best COI policy of any medical journal, even though it limits financial conflicts to only the last three years [25]. It mentions "personal relationships or rivalries, academic competition, or intellectual beliefs" and states that "the editor may use such information as a basis for editorial decisions, and will publish such disclosures if they are believed to be important to readers in judging the manuscript." This grants great discretion to the editors of the journal, who could themselves be subject to bias when assessing COI.

Let's face it: anyone who conducts scientific work can, and most likely will, be biased in some way or another. Authors who feel tempted not to disclose competing interests should be clear about the fact that readers continue to act as "peer reviewers" after publication, increasing peer review by orders of magnitude [26]. The penalties for non-disclosure of relevant conflicts must also be severe (see below).

## Rule 4: Trials Must Be Reported Accurately, As Well As Registered

Another important issue concerns clinical trials. While some countries now mandate registration of clinical trials on public registries like clinicaltrials.gov, there are still many that do not, and even those that do still allow exceptions for certain circumstances and first-in-human (FIH) trials. If research integrity is to be taken seriously, there should be no exceptions. Furthermore, despite the emphasis on registration, reporting of results is actually much more important, and registration is only a means to this end [27]. As such, requiring registration without requiring reporting is pointless. Sanctions must be introduced for those not publishing results or publishing only partial results. The United Kingdom National Health Service (NHS) Research Ethics Service recently made trial registration a condition of approval [28], and the European Union's new clinical trial regulation mandates summaries of results on a new trial registry [29]. These are important steps in the right direction, but full and accurate disclosure of results is necessary to ensure research integrity. Commendably, the Public Library of Science (PLOS) journals now require publication of raw data alongside research articles and adherence to reporting guidelines like CONSORT [30]. Further investment of resources in ethics committees must be sufficient to ensure monitoring of registration and publication of results, and all new projects should only be approved subject to agreements concerning publication and analysis. For example, the Avon Longitudinal Study of Parents and Children requires all projects to guarantee "the right to check that all objectives in the original proposal are completed by cross reference to publications and make any additional analyses that were in the initial proposal but that have not been published via letters to journals and/or on our website, in order to avoid publication bias [31]."

## Rule 5: Any Use of Metrics in Research Decisions Must Be Evidence Based

Another area of research integrity that is often neglected concerns funding and employment decisions. Misuse of impact factors and other metrics means that funding and employment are based on irrelevant measures, which compromises the research process [15]. Any metric-based decisions must be demonstrably evidence-based, possibly with an emphasis on the personal citations of researchers, rather than on impact factors of journals. Misallocation of research funding and misidentification of research priorities can mean that the wrong questions are being asked and the wrong people are being employed as researchers, which also threatens the integrity of the research process [32].

## Rule 6: All Breaches of Integrity Guidelines Should Be Punished or Remediated

An indicator of the way in which low-level misconduct is tolerated is the lack of sanctions for those who breach guidelines. It is hardly surprising that there is little punishment when there is very little detection or enforcement of things like dishonest authorship attribution. In order to change this, palpable sanctions must be introduced. For example, if a journal discovers that a COI was not disclosed by an author, the paper in question could be withdrawn in the same way as if data distortions had been discovered [33,34]. If a researcher is found to have added guest authorship in exchange for some favour, he or she should face disciplinary proceedings without exception. And if a researcher is found to have suppressed or manipulated results, the result of any such proceedings should be dismissal. (In the case of the pharmaceutical industry, in which unethical manipulation of data has been demonstrated, fines for misconduct should be vastly increased.) The financial costs of properly enforcing research integrity policies may be quite high [35], but the moral and scientific cost of failing to act would be greater. The "pyramid of sanctions" approach, with escalating levels of penalties, is likely to be appropriate here [36].

## Rule 7: All Institutions Must Have Clear Procedures for Raising Concerns and Protections for Those Who Do So

A lack of rigor is evident throughout current integrity systems. Assuming that they are aware of the principles of research integrity (see Rule 10), junior researchers, in particular, often have no clear path by which to raise concerns. If such a pathway exists, they often risk losing their jobs or their careers because no protections exist for whistleblowers; this will make many decide not to raise concerns. Even if they are brave enough to do so, there will probably not be defined sanctions that can be brought against the offender. And even if there are, research integrity officers (RIOs) often wield very little power, so there may not be any consequences for the unethical senior researcher, even if there are negative consequences for the ethical junior researcher. All universities and research institutions must have clear pathways for raising concerns, protections for whistleblowers, defined punishments for wrongdoers, and strong powers for RIOs.

## Rule 8: Raising Concerns about Suspected Misconduct Must Be Mandatory

All integrity policies should mandate whistleblowing on integrity issues. To facilitate this, all institutions should provide career protection for those who raise concerns about their superiors. Only if whistleblowers are truly protected will research integrity be fostered.

## Rule 9: RIOs Must Have the Power to Enforce Integrity Policies

Even if integrity guidelines are strengthened in line with these suggestions, strong rules are meaningless without the means to enforce them. As well as making integrity policies stronger and more visible, RIOs at universities, companies, and other research institutions must be given the muscle required to punish those in breach. In practice, this means that RIOs must have the power to initiate academic misconduct proceedings themselves, without explicit approval of deans or others in the university hierarchy.

## Rule 10: Integrity Policies Must Be Highly Visible and Understood

We have assumed so far that researchers are aware both of the basic principles of research integrity and of any relevant institutional or national guidelines. While most researchers probably do have some idea of what constitutes misconduct, very many are unaware of their university's specific policies [37]. Students and staff must be made aware of the standards they are expected to meet; there is no point in having a zero-tolerance approach to scientific misconduct if no one knows that the standard of research integrity has been set so high. As well as making their own integrity guidelines highly visible, institutions should make integrity education a key part of undergraduate and postgraduate courses across all disciplines.

Clearly, our list of rules is not exhaustive, and other requirements may need to be added. Some will object, claiming that it is not necessary to set the bar so high. However, many people will always do the minimum required in terms of integrity. Therefore, the minimum must be set very high in order to protect research integrity. Some will also argue that it is impractical to set the bar so high for such "small" breaches. But if integrity is truly important, researchers must be vigilant for any and all cases of misconduct, not just the really obvious, really bad cases. The various national and international guidelines [38] must be more specific and better enforced in order to prevent them from being used as a camouflage for misconduct.

## Acknowledgments

TE contributed his work to this manuscript whilst he was on sabbatical at the University of Oxford, Sleep and Circadian Neuroscience Institute (SCNi), and enjoyed stimulating working conditions at Brasenose College.

## Author Contributions

DS had the original idea for the article and wrote the first draft. TE revised and improved the article through substantial intellectual contributions, including improving the evidence base for the arguments. Both authors approved the final version.

## References

1. Godecharle S, Nemery B, Dierickx K (2013) Guidance on research integrity: no union in Europe. Lancet 381:1097–1098. doi: 10.1016/S0140-6736(13)60759-X PMID: 23540852
2. Wager E (2013) The UK should lead the way on research integrity. BMJ 346:f2348. doi: 10.1136/bmj.f2348 PMID: 23610377
3. Lancet (2002) The next step: ensuring integrity of scientific research. Lancet: 360:499. PMID: 12241647
4. Lancet (2012) Promoting research integrity: a new global effort. Lancet 380: 1445. doi: 10.1016/S0140-6736(12)61822-4 PMID: 23101698
5. Titus SL, Wells JA, Rhoades LJ (2008) Repairing research integrity. Nature 453:980–982. doi: 10.1038/453980a PMID: 18563131
6. Martinson BC, Anderson MS, de Vries R (2005) Scientists behaving badly. Nature 435: 737–738. PMID: 15944677
7. Foundation/ALLEA ES. European code of conduct for research integrity. http://www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf. Accessed: March 4, 2015.
8. Council ER. (2012) European scientific misconduct strategy. http://erc.europa.eu/sites/default/files/document/file/ERC_Scientific_misconduct_strategy.pdf. Accessed: March 4, 2015.
9. Associated Press in Tokyo (2014) Stap cells: research paper on stem cell breakthrough was partly falsified. The Guardian. http://www.theguardian.com/science/2014/apr/01/stap-cells-research-paper-on-stem-cell-breakthrough-was-partly-falsified. Accessed: March 4, 2015.
10. Jefferson T, Doshi P (2014) Multisystem failure: the story of anti-influenza drugs. BMJ 348: g2263. doi: 10.1136/bmj.g2263 PMID: 24721793
11. van Raaij MJ (2010) Guest authors: for contributors only. Nature 468: 765. doi: 10.1038/468765d PMID: 21150982
12. Rennie D, Yank V, Emanuel L (1997) When authorship fails. A proposal to make contributors accountable. JAMA 278: 579–585. PMID: 9268280
13. ICMJE (2014) Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated December 2014. http://www.icmje.org/recommendations. Accessed: March 4, 2015.

14. Shaw D (2011) The ICMJE's definition of authorship is illogical and unethical. BMJ 343: d7192. doi: 10.1136/bmj.d7192 PMID: 22065677
15. Shaw D (2014) The prisoners' dilemmas: Authorship guidelines and impact factors: between a rock and a hard place. EMBO reports 15: 635–637. doi: 10.1002/embr.201338348 PMID: 24781693
16. Smith R (1997) Authorship is dying: long live contributorship. BMJ 315: 696. PMID: 9314747
17. Kaplan K (2010) Academia: The changing face of tenure. Nature 468: 123–125. PMID: 21157983
18. Glass RM (2010) Guest authors: no place in any journal. Nature 468: 765. doi: 10.1038/468765e PMID: 21150979
19. BMJ. Authorship and contributorship. http://www.bmj.com/about-bmj/resources-authors/article-submission/authorship-contributorship. Accessed: March 4, 2015.
20. Yank V, Rennie D (1997) Disclosure of researcher contributions: a study of original research articles in The Lancet. Annals of internal medicine 130: 661–670.
21. Rochon et al. Financial Conflicts of Interest Checklist 2010 for clinical research studies. Open Medicine 2010; 4(1):E69–E91. PMID: 21686297
22. Goldacre B (2013) Bad Pharma. UK: Fourth Estate.
23. Horrobin DF (1999) Beyond conflict of interest. Non-financial conflicts of interest are more serious than financial conflicts. BMJ 318: 466. PMID: 10084844
24. Shaw DM (2014) A piece of my mind. Beyond conflicts of interest: disclosing medical biases. JAMA 312: 697–698. doi: 10.1001/jama.2014.8035 PMID: 25138329
25. The Lancet. Statements, permissions, and signatures. http://www.thelancet.com/lancet-information-for-authors/statements-permissions-signatures. Accessed: March 4, 2015.
26. Erren TC (2012) Competing interests: Judged in perpetuity. Nature 488: 590. doi: 10.1038/488590c PMID: 22932368
27. Chalmers I, Glasziou P, Godlee F (2013) All trials must be registered and the results published. BMJ 346: f105. doi: 10.1136/bmj.f105 PMID: 23303893
28. Health Research Authority. Transparency, registration and publication. http://www.hra.nhs.uk/resources/during-and-after-your-study/transparency-registration-and-publication/. Accessed: March 4, 2015.
29. All Trials. (2013) Agreement reached in Europe on Clinical Trials Regulation. http://www.alltrials.net/2013/agreement-reached-europe-clinical-trials-regulation. Accessed: March 4, 2015.
30. Bloom T (2013) Data Access for the Open Access Literature: PLOS's Data Policy. http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/. Accessed: March 4, 2015.
31. Avon Longitudinal Study of Parents and Children. Access policy. v. 6.1 (June 2015) http://www.bristol.ac.uk/media-library/sites/alspac/documents/Access%20Policy.pdf (Accessed 13/9/15)
32. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, et al. (2014) How to increase value and reduce waste when research priorities are set. Lancet 383: 156–165. doi: 10.1016/S0140-6736(13)62229-1 PMID: 24411644
33. Fanelli D (2013) Why growing retractions are (mostly) a good sign. PLoS medicine 10: e1001563. doi: 10.1371/journal.pmed.1001563 PMID: 24311988
34. Fanelli D (2014) Publishing: Rise in retractions is a signal of integrity. Nature 509: 33. doi: 10.1038/509033a PMID: 24784207
35. Michalek AM, Hutson AD, Wicher CP, Trump DL (2010) The costs and underappreciated consequences of research misconduct: a case study. PLoS medicine 7: e1000318. doi: 10.1371/journal.pmed.1000318 PMID: 20808955
36. Ayers I and Braithwaite J. Responsive Regulation. Transcending the deregulation debate. New York: Oxford University Press, 1992.
37. Boyd EA1, Cho MK, Bero LA. Financial conflict-of-interest policies in clinical research: issues for clinical investigators. Acad Med. 2003 Aug; 78(8):769–74. PMID: 12915362
38. Resnik DB, Master Z (2013) Policies and initiatives aimed at addressing research misconduct in high-income countries. PLoS Medicine 10:e1001406. doi: 10.1371/journal.pmed.1001406 PMID: 23555198

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for a Computational Biologist's Laboratory Notebook

**Santiago Schnell**[1,2,3]*

**1** Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, **2** Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, **3** Brehm Center for Diabetes Research, University of Michigan Medical School, Ann Arbor, Michigan, United States of America

* schnells@umich.edu



OPEN ACCESS

**Citation:** Schnell S (2015) Ten Simple Rules for a Computational Biologist's Laboratory Notebook. PLoS Comput Biol 11(9): e1004385. doi:10.1371/journal.pcbi.1004385

**Editor:** Scott Markel, Accelrys, UNITED STATES

**Published:** September 10, 2015



**Funding:** This work is supported by National Institute of Diabetes and Digestive and Kidney Diseases grant R25 DK088752. Additional funding is provided by the University of Michigan Protein Folding Disease Initiative. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

One of the major hurdles I face as the head of a computational biology laboratory is convincing my research team—particularly those pursuing exclusively mathematical and computational modeling—that they need to keep a laboratory notebook. There seems to be a misconception in the computational biology community that a lab notebook is only useful for recording experimental protocols and their results. A lab notebook is much more than that. It is an organizational tool and memory aid, which serves as the primary record of scientific research and activity for all scientists. It also serves as a legal record of ownership of the ideas and results obtained by a scientist. Here, I present the best practices (summarized as ten rules) for keeping a lab notebook in computational biology, for scientists pursuing exclusively "dry" research.

## Rule 1: Learn Your Institution's or Laboratory's Notebook Policy

A lab notebook is an important tool for good record-keeping, research management, to protect intellectual property and prevent fraud [1]. Leading research institutions, research and development divisions in companies, and universities have comprehensive lab notebook policies, which research laboratories should implement. In the absence of an institutional policy, your research group should have a policy to explain to all team members the process for daily record-keeping and maintaining laboratory records. If your institution or laboratory does not have a standard policy, the following rules provide you with some guidelines for keeping a record of your scientific activities—a record that will likely be very important for you, your research supervisor, or peers.

## Rule 2: Select the Right Medium for Your Lab Notebook

There are three types of lab notebooks: the bound or stitched notebook, the loose-leaf or three-ring binder notebook, and the computer-based electronic notebook [1]. Each of these notebooks has its advantages and disadvantages [2]. You will need to select the right type of notebook for your research. Most computational biologists work on several projects at the same time. If you find it too complicated to keep all of your projects in a single lab notebook, you can maintain a lab notebook for each project. Alternatively, you can use a ring binder, in which each project is maintained behind a separate tab divider inside the binder. You also have the option of using an electronic lab notebook. The advantage of electronic notes is that they can be searched easily [3], and computer-generated figures can be quickly copied to the notebook. If you do not identify the right technology, it can be very time-consuming to make a polished electronic notebook. Some laboratories [4] are writing lab notebooks using Microsoft

Word, saving as "Web Page," and automatically transfer the entries into a WordPress.com blog. Microsoft OneNote is a proprietary software option. If you write a lot of computer code and collect large datasets, the electronic notebook will be a better option for you because you can store and link your code and data to your electronic lab records [5]. Electronic lab notebooks can also be shared and accessed easily online with collaborators and lab members [3,4]. However, if you have not identified the right technology for an electronic lab notebook, then you should organize your computer code and data in a safe medium (e.g., hard drive, CD-ROM, cloud storage, version-control databases, paper copies) [6] and record in your notebook where the code or data can be found.

## Rule 3: Make the Habit of Keeping the Lab Notebook in Your Desk

You need to keep your lab notebook at hand, and write things down while you are working. If you rely on your recollection to remember a good idea, a suggestion during a meeting, or an important step in your data analysis or model simulation, you can find yourself in a situation where you no longer remember that critical thought. Furthermore, you will find that writing provides you the opportunity to reflect on these ideas as you put together a logical argument supporting your conjectures, results, or conclusions.

## Rule 4: Record All Scientific Activities in Your Lab Notebook

There are scientists who believe that lab notebook records should be limited to wet or dry lab experimental entries. However, the intellectual activity of a theoretical scientist is not limited to experiments to test hypotheses. Thinking about the possible directions of your research and theorizing about how a system works is often how scientific breakthroughs are made. If you use paper pads or other assorted pieces of paper to write down your ideas or take notes during meetings and seminars, you may lose important items or waste too much time looking for ideas later. Recording your scientific activities will solve this problem. Scientists should record every experiment, every result, every research meeting, notes from seminars, research conference calls, thoughts related to their research problem—all of these items go into their lab notebooks. In a very real way, the lab notebook is a chronological log of everything scholarly a scientist does. Each lab notebook entry should be written immediately after the activity or work was performed.

## Rule 5: Every Entry Should Be Recorded with a Date, Subject, and Protocol

The most logical organization of a lab notebook is chronological. Each entry should contain the date it was made and subject of the entry. If you make distinct sets of entries in the same day, you should separate them by using heading titles and leave sufficient space between the entries [1]. The titles of your entries are important. They should be short, sharp, and informative, as you will use them to build a table of contents for your lab notebook. If you are using a paper lab notebook, you will need to write the date and time stamp and make your entries legible, written in permanent ink and in a language accessible to everyone in the laboratory. If you use an electronic lab notebook, the date and time stamps will be entered automatically for each new subject.

You should also include a brief background for each entry [2]. It could be a short synopsis of your thought process that explains why this entry is important for your scientific work. You can support it with references published in the literature, ideas learned from a research talk, or a previous model developed by your laboratory. Then, record everything you do for this specific entry. If you make a mistake in your paper lab notebook, put a thin line through the mistake and write the new information next to it [1]. Never erase or destroy an entry, because errors or mistakes are an important part of the scientific process.

## Rule 6: Keep a Record of How Every Result Was Produced

The gold standard of science is reproducibility [7]. You need to keep a record of how every result was produced in your in silico experiments, statistical analyses, and mathematical or computational models. Noting the sequence of steps taken allows for a result or analysis to be reproduced. For every step in your model, analysis, or experiment, you should record every detail that will influence its execution [8]. This includes the preparation of your wet or dry experiment, preprocessing, execution with intermediate steps, analysis, and postprocessing of results [1,2]. You should also store the raw data for every figure. This will allow you to have the exact values for the visualization of your results. It will also give you the opportunity to redraw figures to improve their quality or ensure visual consistency for publication.

If a result is obtained through a mathematical model, algorithm, or computer program, you need to include the equations and name and version of the algorithm or computer program, as well as the initial conditions and parameters used in the model. In many instances, a statistical or computational analysis creates intermediate data [9]. You should record intermediate results because they can reveal discrepancies with your final results, particularly if your analysis is time consuming or readily executable. At the same time, they can help you track any inconsistencies in your analysis or algorithms [8]. Electronic lab notebooks can be very convenient for storing data and linking to computer mathematical models, algorithms, and software stored in cloud mass storage systems.

## Rule 7: Use Version Control for Models, Algorithms, and Computer Code

As a mathematical and computational biologist, you will be updating your models, algorithms, or computer programs frequently. You will also create scripts containing initial conditions and parameters to run analyses or simulations. Changes in your models, algorithms, programs, or scripts could drastically change your results. If you do not systematically archive changes, it will be very difficult or impossible to track the codes that you used to generate certain results [8,9]. Nowadays, there are version control systems to track the evolution of algorithms and computer programs or changes in scripts. Bitbucket, Git, Subversion, and Mercurial are among the most widely used version-control systems. You should use a standardized name system to identify changes. If you have a paper lab notebook, you should record the name and location of the scripts. Those using electronic lab notebooks can add links to each version of their scripts or programs.

## Rule 8: Keep a Lab Notebook That Can Serve As a Legal Record of Your Work

The lab notebook serves as a legal record of ownership of ideas and results [10]. Lab notebooks can serve to determine authorship in scientific papers or rights for establishing copyright or patent rights. If you do not feel comfortable walking around with a notebook or having more than one lab notebook, you can still record all your notes in paper pads, but you should file them in a ring binder using the indexing system of a lab notebook. You can also use a tablet if you keep an electronic lab notebook. However, this is not generally advisable. At the moment, bound notebooks with numbered pages are the only legally recognized option to record and protect your work. Electronic records can be printed out on a regular basis and then bound to form a legally recognized laboratory notebook. If you keep a loose-leaf lab notebook, you should have a parallel hardbound notebook summarizing progress on your projects as a legal record of your work. For each lab notebook entry, clearly indicate who did what work and who was present for a discussion or in silico experiment; this is particularly important for collaborative projects. In addition, each entry should be signed by you and cosigned by a coworker or supervisor. Otherwise, the entry will not serve as a legally valid record.

It is important to note that lab notebooks can be subpoenaed as evidence in cases of scientific misconduct by your institution or governmental funding agencies. According to educational material provided by the Office of Research Integrity at the United States Department of Health and Human Services and the University of New Hampshire, "The integrity of research and scholarly activities depends on accurate, detailed, organized, complete, and accessible data" [11]. From the legal point of view, you keep a lab notebook to reflect your own scientific integrity and to give others the opportunity to corroborate the results of your research. If you keep an electronic lab notebook, you can make it open to the public. This will make your work more transparent to the scientific community, and it will give the opportunity for other scientists to learn more about your research, including the experiments and models that did not work out.

## Rule 9: Create a Table of Contents for Your Lab Notebook

You should record the titles of all entries in your lab notebook in a table of contents as you finish each entry or day. The idea of this index is to help you, your research supervisor, or someone else find the record of your scientific work efficiently [1,2]. There are multiple formats for the table of contents. You should use the format agreed upon in your laboratory. It is generally advisable that each entry in the table of contents has the date the entry has been made, the subject of your entry, and where in the lab notebook the entry can be found. To find information easily in your lab notebook, you should always start an event in a new page. Then, label the event accordingly ("Research Review with Dr. Williams," "Seminar by Dr. Murray," "Thoughts on Project X," etc.) and date it. Once you have done this, you can start taking notes and take as many pages as you need. If you have the standard paper lab notebook, the pages will be numbered. Now you will be in the position to move forward with the critical step: create a table of contents on the first page of your notebook, where you will log the event and page number. If you do not have the standard lab notebook, you can put a page number at the top of each pair of pages counting by two. An advantage of electronic lab notebooks is that you do not need to worry about creating a table of contents because it will be done automatically for you.

## Rule 10: Protect Your Lab Notebook

As your research activities are funded by or through your academic institution, your lab notebook does not belong to you; it belongs to your institution [1,10]. Your lab notebook is part of the scientific legacy of your laboratory. Therefore, you need to protect your lab notebook. Paper lab notebooks should not be taken home. When you leave the laboratory each day, you should leave your lab notebook in a location where your research supervisor can find it. Ideally, you should lock it in the same place every day. If your research institution or lab supervisor allows, you will be able to make copies of your lab notebook, but the original belongs to the institution that paid your salary and handled your research funds.

Laboratories, research institutes, and universities have archival systems in operation to store lab notebooks. You might find this surprising, but lab notebooks generally end up in your university or research institution's library. Maybe one day you will become a very important scientist, and your lab notebook will be shown in an exhibit, like Charles Darwin's Notebook B, sketching the relationships between related species, or Albert Einstein's notebook from 1928, elaborating how the Einstein field equations (a set of ten equations in his theory of general relativity) can be extended to explain electricity and magnetism.

## References


1. Thomson JA (2007). How to Start—and Keep—a Laboratory Notebook: Policy and Practical Guidelines. In: *Intellectual Property Management in Health and Agricultural Innovation*: *A Handbook of Best Practices* (eds. Krattiger A, Mahoney RT, Nelsen L, et al.). MIHR: Oxford, U.K., pp. 763–771.
2. National Institutes of Health, Office of Intramural Training and Education's Webinar on Keeping a Lab Notebook. https://www.training.nih.gov/OITEtutorials/OITENotebook/Notebook.html
3. Sandefur CI (2014). Blogging for electronic record keeping and collaborative research. Am J Physiol Gastrointest Liver Physiol. 307:G1145–G1146. doi: 10.1152/ajpgi.00384.2014 PMID: 25359540
4. De Polavieja Lab. Open LabBook. http://www.neural-circuits.org/open-labbook
5. Todoroki S, Konishi T, Inoue S (2006). Blog-based research notebook: personal informatics workbench for high-throughput experimentation. Appl Surf Sci 252: 2640–2645.
6. Ruggiero P, Heckathorn MA (2012). Data backup options. Department of Homeland Security, United States Computer Readiness Plan. https://www.us-cert.gov/sites/default/files/publications/data_backup_options.pdf.
7. Jasny BR, Chin G, Chong L, Vignieri S (2011). Data replication & reproducibility. Again, and again, and again .... Introduction. Science. 334(6060):1225. doi: 10.1126/science.334.6060.1225 PMID: 22144612
8. Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013). Ten simple rules for reproducible computational research. PLoS Comput Biol. 9(10):e1003285. doi: 10.1371/journal.pcbi.1003285 PMID: 24204232
9. Peng RD (2011). Reproducible research in computational science. Science. 334:1226–7. doi: 10.1126/science.1213847 PMID: 22144613
10. U.S. Department of Health & Human Services, Office of Research Integrity. Notebook and data management. http://ori.hhs.gov/education/products/wsu/data.html
11. U.S. Department of Health & Human Services, Office of Research Integrity. Data Management Responsibilities. http://ori.hhs.gov/education/products/unh_round1/www.unh.edu/rcr/DataMgt-Responsibilities01.htm

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Reducing Overoptimistic Reporting in Methodological Computational Research

**Anne-Laure Boulesteix***

Institute for Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany

* boulesteix@ibe.med.uni-muenchen.de



**Citation:** Boulesteix A-L (2015) Ten Simple Rules for Reducing Overoptimistic Reporting in Methodological Computational Research. PLoS Comput Biol 11(4): e1004191. doi:10.1371/journal.pcbi.1004191

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** April 23, 2015



**Funding:** ALB is supported by grant BO3139/2-2 from the German Science Foundation (DFG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

In most scientific fields, and in biomedical research in particular, there have long been many discussions on how to improve research practices and methods. The trend has increased in recent years, as illustrated by the series on "reducing waste," published in The Lancet in January 2014 [1], or by the recent essay by John Ioannidis on how to make published results more true [2], which echoes his earlier provocative paper entitled "Why most published research findings are false" [3]. One of the important aspects underlying these discussions is that biomedical literature is most often overoptimistic with respect to, for example, the superiority of a new therapy or the strength of association between a risk factor and an outcome. Published results appear more significant, more spectacular, or sometimes more intuitive—in a word, more "satisfactory"—to authors and readers than they actually would if they reflected the truth. Causes of this problem are diverse, numerous, and interrelated. The effects of "fishing for significance" strategies or selective/incomplete reporting are exacerbated by design issues (e.g., small sample sizes, many investigated features) [3] or publication bias [4], to cite only a few of the factors at work.

Research and guidelines on how to reduce overoptimistic reporting in the context of computational research, including computational biology as an important special case, however, are surprisingly scarce. Many methodological articles published in computational literature report the (vastly) superior performance of new methods [5], too often in general terms and—directly or indirectly—implying that the presented positive results are generalizable to other settings. Such overoptimistic reporting confuses readers, makes literature less credible and more difficult to interpret, and might even ultimately lead to a waste of resources in some cases. Here I take advantage of the popular "ten-simple-rules" format [6] to address the problem of overoptimistic reporting in methodological computational biology research, that is papers—termed "methodological papers" here—devoted primarily to the development and testing of new computational methods (intended to be used by other researchers on other data in the future) rather than to the biological question itself or the specific dataset at hand.

## Rule 1: Assess the New Method

If your goal is to present a new method and convince readers to use it, assess this new method. Applying it to data to answer a biological question and obtaining plausible, interesting results is nice. But this is not sufficient to establish that the new method has advantages over existing methods, nor is it adequate in providing trustworthy biological results—since the validity of the computational method has not yet been assessed. It is not impossible—but it is difficult—to both assess a new computational method and address a relevant biological question in the same article. The assessment of the new method may be performed in different ways depending on the context, for example, by conducting simulations, applying the method to several real datasets, checking the underlying assumptions in practical examples, etc. Notably, if there exist competing methods for performing the same task, they should be compared to the new method; see Rule 2 for more details.

## Rule 2: Compare the New Method to the Best

A new method will be useful in practice only if it performs better (see Rule 6 for a discussion of "better performance") when compared to the best existing methods performing the same task. The new method should not be compared to old methods no longer in use, to obsolete versions of currently used methods, or to good methods with suboptimal parameter settings: comparing the new method to suboptimal competitors will inevitably make it look better than it actually is. This rule is especially important for research topics common in the literature. For example, for supervised classification based on high-dimensional omics data, tens or even hundreds of methods have already been proposed: a new method should not be solely compared to basic methods such as naive Bayes. See [7] for a more in-depth discussion of this problem. Note, however, that recent methods are not always publicly available as user-friendly computer programs, which may make comparison challenging in practice.

## Rule 3: Consider Enough Datasets

To establish that a new method works well in practice, it is important to evaluate its performance using several datasets, just as it is important to evaluate the efficiency of a new drug based on several patients before recommending it for use on other patients [5,8]. With this analogy in mind, it becomes clear that many datasets are needed if one wants to firmly establish the superiority of a new method, and that the question of "how many" is essentially a statistical question [9].

For example, if one compares the performances of two methods—as measured by a normally distributed criterion—on ten datasets, a paired t-test may be used for statistical inference. Beyond the t-test itself, one may also, for example, derive a confidence interval for the difference between the performances of the two methods, apply a multiple testing procedure if more than two methods are compared, or compute the power of the paired t-test to detect a given difference considered relevant by the researcher [9].

In simulations, it will generally not be a problem to consider a (very) large number of datasets, except in cases where the analysis of each dataset is extremely computationally expensive. If one could generate and analyse infinitely many datasets for a given simulation setting, there would be no need to perform a test to assess the difference between the performances of the considered methods: the distribution of this difference would be known. In practice, one should generate and analyse as many datasets as computationally feasible.

For comparisons based on real datasets, however, it may be difficult to find—and have access to—enough adequate example datasets. For topics such as supervised classification based on high-dimensional omics data, numerous well-documented datasets can be found in publicly available databases like ArrayExpress, GEO, and TCGA—to cite only a few. For more complex or recent research questions or data types, however, it may be difficult to apply the new method to more than one or two illustrative datasets.

If the data examples are merely meant as illustrations, which is also fine, it should be stated clearly that they are not intended to be representative of what would happen with similar datasets [8]: in this situation, interpretations of and conclusions on the performance of the new method should be formulated cautiously.

## Rule 4: Do Not "Fish" for Datasets

Example datasets should not be selected just because they yield favorable results for the new method. Similarly, one should not exclude a dataset from the analysis just because it yields bad results. The dramatic consequences in terms of overoptimism of such a "fishing for datasets" strategy have been assessed elsewhere through theoretical modeling and simulations [10] and empirical studies [7]. Ideally, one should define "inclusion criteria" for datasets (e.g., datasets with a particular size or format, on pre-specified diseases, etc.), apply the methods, and report all results.

These inclusion criteria should reflect the intended field of application: if most real datasets have certain features, then the datasets to be selected should also have this feature. For example, it would be unsuitable to include only large datasets in the study if most datasets in the target research field are smaller; such a study may even produce misleading results, since the relative performances of the considered methods may, to some extent, depend on the dataset's size.

## Rule 5: Think of the No-Free-Lunch Theorem and Report Limitations

No reasonable researcher requires your method to always work better than existing methods. Think of the widely acknowledged "no-free-lunch theorem" [11]. Methods are not characterized by a single criterion; see also Rule 6. Datasets are extremely diverse, and so are the performances of methods when applied to them. Referees are supposed to be reasonable researchers, so they will most likely not prevent the publication of your paper simply because your new method is not perfect in all situations and in all respects. Do not forget that, and interpret your results accordingly.

In particular, report limitations of your method and study. In medical literature—those reports on new medical discoveries obtained with an existing data analysis method described elsewhere—the section on "limitations of the study" is considered crucial. Limitations of the method's applicability, practical problems, implementation issues, and pitfalls related to the study design should also be stated clearly in a methodological paper. This rule is related to Philip Bourne's Rule 2 on objectivity in the first ten-rules article [12].

## Rule 6: Consider Several Criteria

Do not become obsessed by a single objective performance criterion, such as, in the case of supervised learning, predictor error. Many other aspects of a new method are important, for example, its computational efficiency, its generalizability, its conceptual simplicity, its lack of sensitivity to the choice of parameters or starting values, and its robustness against the violation of assumptions, to cite only a few. Note that, in computational biology, the ground truth is often unknown in real data applications, which makes the measurement of performance difficult. In these situations, the ability of the new method to uncover the truth can be evaluated using simulations (see Rule 8), and alternative criteria, such as those listed above, can be used to assess the method's behaviour in real data settings. Considering several criteria naturally reduces overoptimistic reporting because—most often—no method is better with respect to all criteria. Further, such considerations also provide a more complete picture of the method's performance and utility.

## Rule 7: Validate Using Independent Data

The new method should be evaluated using data that were not used during the development phase. For example, consider the case of a new machine learning method for supervised classification, such as a new variant of support vector machines (SVM). Its prediction error on a real dataset of moderate size is typically measured through cross-validation (CV) techniques. One obtains as many cross-validation estimates of predictor error as considered datasets. Now imagine modifying some of the new method's characteristics in a trial-and-error process, gradually improving these CV estimates [13]. Ultimately, the CV error estimates could be relatively small as a result of this optimization process and the new method would seemingly work well if evaluated using these datasets: the new method would overfit the datasets used for its development. But this says nothing about the ability of the new method to work well on other, independent datasets—for which the prediction error may be much higher. See also [7] for an empirical study on the potential impact of such optimization mechanisms in practice. For proper evaluation, one has to use other—independent—datasets, which are not examined by the researchers until the new method is fully specified.

Note that this problem is similar to the well-known problem that in machine learning one should not evaluate a prediction rule on the training data on which it was fit. However, here we are concerned with the validation of the methods' general performance rather than with the validation of results obtained with these methods on a specific dataset. In our example, we consider the evaluation of the general performance of the SVM variant rather than the evaluation of the prediction rule resulting when this SVM variant is applied to a specific training dataset.

## Rule 8: Design Simulations Appropriately

A simulation should ideally encompass different settings (e.g., different data sizes, different correlation structures, etc.), which roughly reflect the type of data encountered in the intended area of application. Simulations should not be limited to artificial datasets corresponding exactly to the assumptions underlying the new method, as this would obviously favor the new method. Other data types should be considered as well. Ultimately, the practical relevance of a simulation depends on the similarity between the considered simulation settings and the real datasets in the area of application. Finally, while interpreting simulation results, one should not forget that simulated datasets represent but a tiny dot in the infinite space of possible parameters and settings, which can be seen as an intrinsic limitation of simulations—needing to be discussed, as stipulated by Rule 5. In practice, this problem can be stressed, for example, through phrases such as "in our simulation setting, we found that. . ."

## Rule 9: Provide All Information

The new method's definition, its underlying assumptions, its parameters, the study design, data preparation steps and, last, but not least, implementation issues and computer codes for reproducibility purposes [14] should be carefully reported. Whenever possible, data should be made publicly available so that interested readers can rerun analyses, check results, try alternative analysis strategies, or better compare the study's results to that of their own study's. Reporting has been a widely discussed topic in the last few years in biomedical research [15]. We claim that it also deserves attention in the context of methodological computational research. High-quality reporting, including, but not limited to, computational reproducibility through publication of codes and (whenever possible) data, reduces overoptimism and its impact by increasing transparency and allowing readers to better interpret results to counter the potentially overoptimistic statements of the authors.

## Rule 10: Read the Other Ten Simple Rules Articles

Some rules presented in other ten-simple-rules articles are also directly or indirectly related to overoptimistic reporting, for instance, those on writing papers [16], better figures [17], getting published [12], efficient computational research [18], and reproducible research [14].

## Conclusion

Some amount of overoptimism is certainly unavoidable in literature. From a purely statistical point of view, type I error is non-zero even if a test is performed correctly. Correspondingly, one cannot expect literature to be free of false positive research findings.

Of crucial note is that the problem of overoptimism is related to publication policies and publication bias. As long as journal editors and referees reject sound studies on sensible ideas simply because the new method was not vastly superior to existing methods, authors will always have to be somewhat overoptimistic (possibly also including ourselves!). Reducing the so-called publication bias in the context of methodological research is a challenge that still has to be addressed both from an epistemological point of view (what is actually publication bias?) and from a practical/editorial perspective (which reduction measures could be reasonably undertaken by journals?).

To conclude, overoptimistic reporting is a problem with multiple facets. Advice to authors and solutions to reduce overoptimism should go beyond a ten-rules article. However, following the ten simple rules above can have a considerable influence on alleviating the problem of overoptimism in reporting.

## Acknowledgments

I thank Rory Wilson for helpful comments.

## References

1. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP, et al. (2014) Biomedical research: increasing value, reducing waste. The Lancet 383: 101–104. doi: 10.1016/S0140-6736(13)62329-6 PMID: 24411643
2. Ioannidis JP (2014) How to make more published research true. PLoS Medicine 11: e1001747. doi: 10.1371/journal.pmed.1001747 PMID: 25334033
3. Ioannidis JP (2005) Why most published research findings are false. PLoS Medicine 2: e124. PMID: 16060722
4. Easterbrook PJ, Gopalan R, Berlin J, Matthews DR (1991) Publication bias in clinical research. The Lancet 337: 867–872. PMID: 1672966
5. Boulesteix AL, Lauer S, Eugster MJE (2013) A plea for neutral comparison studies in computational sciences. PLoS One 8: e61562. doi: 10.1371/journal.pone.0061562 PMID: 23637855
6. Dashnow H, Lonsdale A, Bourne PE (2014) Ten simple rules for writing a plos ten simple rules article. PLoS Computational Biology 10: e1003858. doi: 10.1371/journal.pcbi.1003858 PMID: 25340653
7. Jelizarow M, Guillemot V, Tenenhaus A, Strimmer K, Boulesteix AL (2010) Over-optimism in bioinformatics: an illustration. Bioinformatics 26: 1990–1998. doi: 10.1093/bioinformatics/btq323 PMID: 20581402
8. Boulesteix AL (2013) On representative and illustrative comparisons with real data in bioinformatics: response to the letter to the editor by Smith et al. Bioinformatics 29: 2664–2666. doi: 10.1093/bioinformatics/btt458 PMID: 23929033
9. Boulesteix AL, Hable R, Lauer S, Eugster MJE. A statistical framework for hypothesis testing in real data comparison studies. The American Statistician [Internet]. 2015 Jan. Available from: 10.1080/00031305.2015.1005128#.VQhuq47F-gY.
10. Yousefi MR, Hua J, Sima C, Dougherty ER (2010) Reporting bias when using real data sets to analyze classification performance. Bioinformatics 26: 68–76. doi: 10.1093/bioinformatics/btp605 PMID: 19846436

11. Wolpert D (2001) The supervised learning no-free-lunch theorems. In: Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications. Citeseer, volume 6, pp. 1–20.
12. Bourne PE (2005) Ten simple rules for getting published. PLoS Computational Biology 1: e57. PMID: 16261197
13. Rocke DM, Ideker T, Troyanskaya O, Quackenbush J, Dopazo J (2009) Papers on normalization, variable selection, classification or clustering of microarray data. Bioinformatics 25: 701–702.
14. Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten simple rules for reproducible computational research. PLoS Computational Biology 9: e1003285. doi: 10.1371/journal.pcbi.1003285 PMID: 24204232
15. Altman DG, Simera I, Hoey J, Moher D, Schulz K (2008) EQUATOR: reporting guidelines for health research. Lancet 371: 1149–1150. doi: 10.1016/S0140-6736(08)60505-X PMID: 18395566
16. Zhang W (2014) Ten simple rules for writing research papers. PLoS Computational Biology 10: e1003453. doi: 10.1371/journal.pcbi.1003453 PMID: 24499936
17. Rougier NP, Droettboom M, Bourne PE (2005) Ten simple rules for better figures. PLoS Computational Biology 10: e1003833. doi: 10.1371/journal.pcbi.1003833 PMID: 25210732
18. Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, Dalchau n, et al. (2014) Ten simple rules for effective computational research. PLoS Computational Biology 10: e1003506. doi: 10.1371/journal.pcbi.1003506 PMID: 24675742

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for the Care and Feeding of Scientific Data

**Alyssa Goodman[1], Alberto Pepe[1]*, Alexander W. Blocker[1], Christine L. Borgman[2], Kyle Cranmer[3], Merce Crosas[1], Rosanne Di Stefano[1], Yolanda Gil[4], Paul Groth[5], Margaret Hedstrom[6], David W. Hogg[3], Vinay Kashyap[1], Ashish Mahabal[7], Aneta Siemiginowska[1], Aleksandra Slavkovic[8]**

**1** Harvard University, Cambridge, Massachusetts, United States of America, **2** University of California, Los Angeles, Los Angeles, California, United States of America, **3** New York University, New York, New York, United States of America, **4** University of Southern California, Los Angeles, Los Angeles, California, United States of America, **5** Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, **6** University of Michigan, Ann Arbor, Michigan, United States of America, **7** California Institute of Technology, Pasadena, California, United States of America, **8** Pennsylvania State University, State College, Pennsylvania, United States of America

## Introduction

In the early 1600s, Galileo Galilei turned a telescope toward Jupiter. In his log book each night, he drew to-scale schematic diagrams of Jupiter and some oddly moving points of light near it. Galileo labeled each drawing with the date. Eventually he used his observations to conclude that the Earth orbits the Sun, just as the four Galilean moons orbit Jupiter. History shows Galileo to be much more than an astronomical hero, though. His clear and careful record keeping and publication style not only let Galileo understand the solar system, they continue to let *anyone* understand *how* Galileo did it. Galileo's notes directly integrated his **data** (drawings of Jupiter and its moons), key **metadata** (timing of each observation, weather, and telescope properties), and **text** (descriptions of methods, analysis, and conclusions). Critically, when Galileo included the information from those notes in *Sidereus Nuncius* [1], this integration of text, data, and metadata was preserved, as shown in Figure 1. Galileo's work advanced the "Scientific Revolution," and his approach to observation and analysis contributed significantly to the shaping of today's modern "scientific method" [2,3].

Today, most research projects are considered complete when a journal article based on the analysis has been written and published. The trouble is, unlike Galileo's report in *Sidereus Nuncius*, the amount of real data and data description in modern publications is almost never sufficient to repeat or even statistically verify a study being presented. Worse, researchers wishing to build upon and extend work presented in the literature often have trouble recovering data associated with an article after it has been published. More often than scientists would like to admit, they cannot even recover the data associated with their own published works.

Complicating the modern situation, the words "data" and "analysis" have a wider variety of definitions today than at the time of Galileo. Theoretical investigations can create large "data" sets through simulations (e.g., The Millennium Simulation Project: http://www.mpa-garching.mpg.de/galform/virgo/millennium/). Large-scale data collection often takes place as a community-wide effort (e.g., The Human Genome project: http://www.genome.gov/10001772), which leads to gigantic online "databases" (organized collections of data). Computers are so essential in simulations, and in the processing of experimental and observational data, that it is also often hard to draw a dividing line between "data" and "analysis" (or "code") when discussing the care and feeding of "data." Sometimes, a copy of the code used to create or process data is so essential to the use of those data that the code should almost be thought of as part of the "metadata" description of the data. Other times, the code used in a scientific study is more separable from the data, but even then, many preservation and sharing principles apply to code just as well as they do to data.

So how do we go about caring for and feeding data? Extra work, no doubt, is associated with nurturing your data, but care up front will save time and increase insight later. Even though a growing number of researchers, especially in large collaborations, know that conducting research with sharing and reuse in mind is essential, it still requires a paradigm shift. Most people are still motivated by piling up publications and by getting to the next one as soon as possible. But, the more we scientists find ourselves wishing we had access to extant but now unfindable data [4], the more we will realize why bad data management is bad for science. How can we improve?

**This article offers a short guide to the steps scientists can take to ensure that their data and associated analyses continue to be of value and to be recognized.** In just the past few years, hundreds of scholarly papers and reports have been written on questions of data sharing, data provenance, research reproducibility, licensing, attribution, privacy, and more—but our goal here is *not* to review that literature. Instead, we present a short guide intended for researchers who want to know why it is important to "care for and feed" data, with some practical advice on how to do that. The final section at the close of this work (Links to Useful Resources) offers links to the types of services referred to throughout the text. **Boldface lettering** below highlights actions one can take to follow the suggested rules.

## Rule 1. Love Your Data, and Help Others Love It, Too

Data management is a repeat-play game. If you take care to make your data

**Citation:** Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, et al. (2014) Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol 10(4): e1003542. doi:10.1371/journal.pcbi.1003542

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

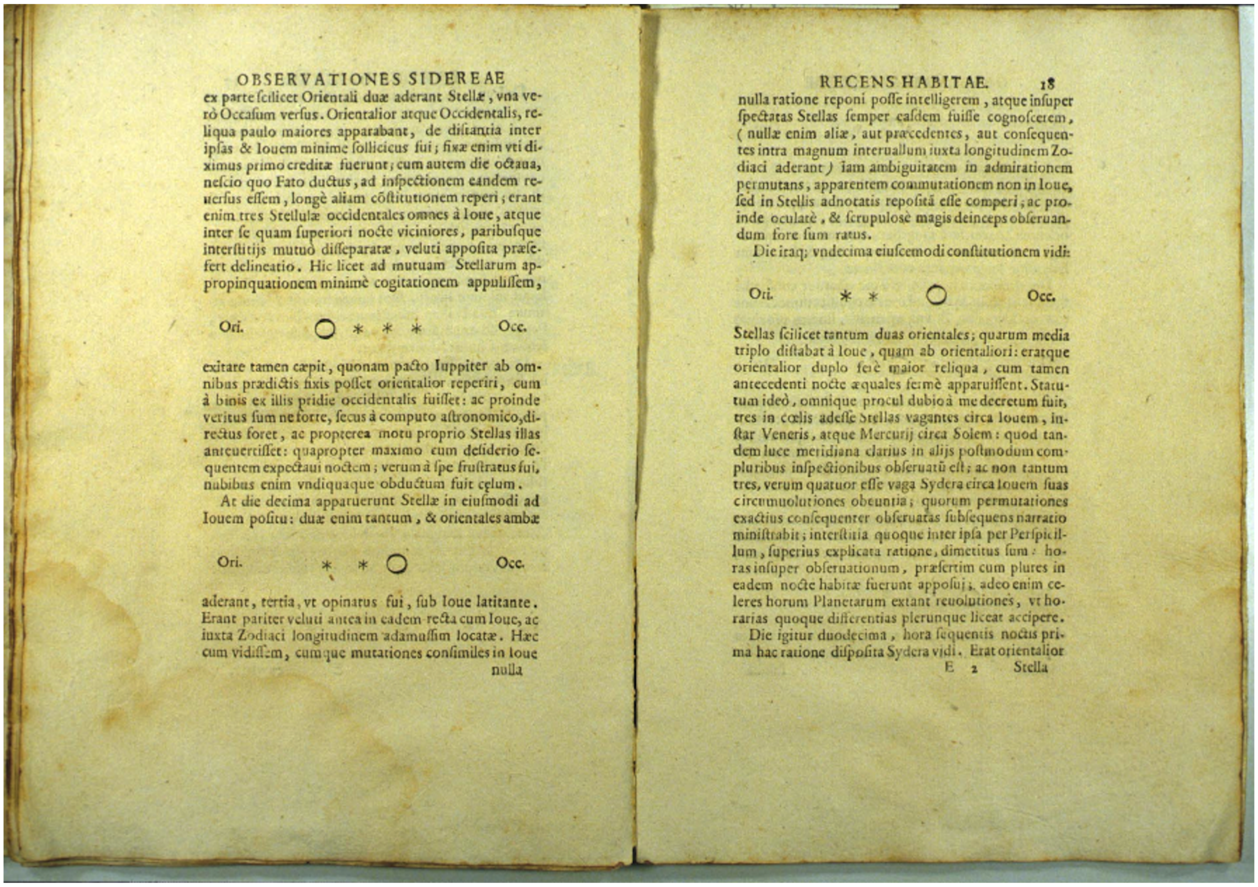
**Published** April 24, 2014



**Funding:** The authors received no specific funding for writing this manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: alberto.pepe@gmail.com



OBSERVATIONES SIDEREAE

ex parte ſcilicet Orientali duæ aderant Stellæ, vna verò Occaſum verſus. Orientalior atque Occidentalis, reliqua paulo maiores apparabant, de diſtantia inter ipſas & Iouem minime ſollicitus fui; fixæ enim vti diximus primo creditæ fuerunt; cum autem die octaua, neſcio quo Fato ductus, ad inſpectionem eandem reuerſus eſſem, longè aliam cõſtitutionem reperi; erant enim tres Stellulæ occidentales omnes à Ioue, atque inter ſe quam ſuperiori nocte viciniores, paribuſque interſtitijs mutuò diſſeparatæ, veluti appoſita præſefert delineatio. Hic licet ad mutuam Stellarum appropinquationem minimè cogitationem appuliſſem,

Ori. Occ.

exitare tamen cæpit, quonam pacto Iuppiter ab omnibus prædictis fixis poſſet orientalior reperiri, cum à binis ex illis pridie occidentalis fuiſſet: ac proinde veritus ſum ne forte, ſecus à computo aſtronomico, directus foret, ac propterea motu proprio Stellas illas anteuertiſſet: quapropter maximo cum deſiderio ſequentem expectaui noctem; verum à ſpe fruſtratus fui, nubibus enim vndiquaque obductum fuit cęlum.

At die decima apparuerunt Stellæ in eiuſmodi ad Iouem poſitu: duæ enim tantum, & orientales ambæ

Ori. Occ.

aderant, tertia, vt opinatus fui, ſub Ioue latitante. Erant pariter veluti antea in eadem recta cum Ioue, ac iuxta Zodiaci longitudinem adamuſſim locatæ. Hæc cum vidiſſem, cumque mutationes conſimiles in Ioue nulla

RECENS HABITAE. 18

nulla ratione reponi poſſe intelligerem, atque inſuper ſpectatas Stellas ſemper eaſdem fuiſſe cognoſcerem, (nullæ enim aliæ, aut præcedentes, aut conſequentes intra magnum interuallum iuxta longitudinem Zodiaci aderant) iam ambiguitatem in admirationem permutans, apparentem commutationem non in Ioue, ſed in Stellis adnotatis repoſitā eſſe comperi; ac proinde oculatè, & ſcrupuloſè magis deinceps obſeruandum fore ſum ratus.

Die itaq; vndecima eiuſcemodi conſtitutionem vidi:

Ori. Occ.

Stellas ſcilicet tantum duas orientales; quarum media triplo diſtabat à Ioue, quam ab orientaliori: eratque orientalior duplo ferè maior reliqua, cum tamen antecedenti nocte æquales fermè apparuiſſent. Statutum ideò, omnique procul dubio à me decretum fuit, tres in cœlis adeſſe Stellas vagantes circa Iouem, inſtar Veneris, atque Mercurij circa Solem: quod tandem luce meridiana clarius in alijs poſtmodum compluribus inſpectionibus obſeruatū eſt; ac non tantum tres, verum quatuor eſſe vaga Sydera circa Iouem ſuas circumuolutiones obeuntia; quorum permutationes exactius conſequenter obſeruatas ſubſequens narratio miniſtrabit; interſtitia quoque inter ipſa per Perſpicillum, ſuperius explicata ratione, dimetitus ſum: horas inſuper obſeruationum, præſertim cum plures in eadem nocte habitæ fuerunt appoſuj; adeo enim celeres horum Planetarum extant reuolutiones, vt horarias quoque differentias plerunque liceat accipere.

Die igitur duodecima, hora ſequentis noctis prima hac ratione diſpoſita Sydera vidi. Erat orientalior

E 2 Stella

**Figure 1. Two pages (scan) from Galilei's *Sidereus Nuncius* ("The Starry Messenger" or "The Herald of the Stars"), Venice, 1610.** On these pages, Galilei combines data (drawings of Jupiter and its moons), key metadata (timing of each observation, weather, and telescope properties), and text (descriptions of methods, analysis, and conclusions).
doi:10.1371/journal.pcbi.1003542.g001

easily available to others, others are more likely to do the same—eventually. While we wait for this new sharing equilibrium to be reached, you can take two important actions. First, cherish, document, and **publish your data**, preferably using the robust methods described in Rule 2. Get started now, as better tools and resources for data management are becoming more numerous, universities and research communities are moving toward bigger investments in data repositories (Rule 8), and more librarians and scientists are learning data management skills (Rule 10). At the very least, loving your own available data will serve *you*: you'll be able to find and reuse your own data if you treat them well. Second, enable and **encourage others to cherish, document, and publish their data**. If you are a research scientist, chances are that not only are you an author, but also a reviewer for a specialized journal or conference venue. As a reviewer, **request that the authors of papers you review provide documentation and access to their data** according to the rules set out in the remainder of this article. While institutional approaches are clearly essential (Rules 8 and 10), changing minds one scientist at a time is effective as well.

## Rule 2. Share Your Data Online, with a Permanent Identifier

Nothing really lasts forever, so "permanent" actually just means long-lasting. For example, your personal web site is unlikely to be a good option for long-term data storage (yet, in the very short run, putting your data on your site is better than doing nothing at all!). In general, although many papers include URLs to give access to datasets, most become inaccessible within a few years [5]. The best option for releasing your data with long-term guarantee is to **deposit them in whatever data archive is the "go to" place for your field**. A proper, trustworthy archive will: (1) assign an identifier such as a "handle" (hdl) or "digital object identifier" (doi); (2) require that you provide adequate documentation and metadata; and (3) manage the "care and feeding" of your data by employing good curation practices. If no such archive exists in your field, there are also generic (non-domain-specific) online services that can host your data and issue persistent identifiers (see Rule 8). Pointers to a few generic repositories are listed in the Links to Useful Resources (section A), and longer compilations of such services are in the Links to Useful Resources (B).

## Rule 3. Conduct Science with a Particular Level of Reuse in Mind

Data from others are hard to use without context describing what the data are and how they were obtained. The W3C Provenance Group (http://www.w3.org/TR/2013/REC-prov-dm-20130430/#dfn-provenance) defines information "provenance" as the sum of all of the processes, people (institutions or agents), and documents (data included!) that were involved in generating or otherwise influencing or delivering a piece of information. Perfect documentation of provenance is rarely, if ever, attained in scientific work today. The higher the quality of provenance information, the higher the chance of enabling data reuse. In general, data reuse is most possible when: 1) data; 2) metadata (information describing the data); and 3) information about the process of generating those data, such as code, are all provided. In trying to follow the rules listed in this article, you will do best if you plan in advance for ways to provide all three kinds of information. **In carrying out your work, consider what level of reuse you realistically expect and plan accordingly.** Do you want your work to be fully *reproducible*? If so, then provenance information is a must (e.g., working pipeline analysis code, a platform to run it on, and verifiable versions of the data). Or do you just want your work to be *inspectable*? If so, then intermediate data products and pseudo-code may be sufficient. Or maybe your goal is that your data is *usable* in a wide range of applications? If so, **consider adopting standard formats and metadata standards early on**. At the very least, **keep careful track of versions of data and code**, with associated dates. Taking these steps as you plan and carry out projects will earn you the thanks of researchers, including you, looking back from the future. (Consult the Links to Useful Resources [E] for a list of tools to package all your research materials with reuse in mind.)

## Rule 4. Publish Workflow as Context

Publishing a description of your processing steps offers essential context for interpreting and reusing data. As such, scientists typically include a "methods" and/or "analysis" section(s) in a scholarly article, used to describe data collection, manipulation, and analysis processes. Computer and information scientists call the combination of the collection methods and analysis processes for a project its "workflow," and they consider the information used and captured in the workflow to be part of the "provenance" of the data. In some cases (mostly in genomics), scientists can use existing workflow software in *running* experiments and in *recording* what was done in those experiments, e.g., Gene Pattern (www.genepattern.org). In that best-case scenario, the workflow software, its version, and settings used can be published alongside data using the other rules laid out here. But, it is rare outside of genomics to see the end-to-end process described in a research paper run, orchestrated, and/or recorded by a single software package. In a plausible utopian future, automated workflow documentation could extend to all fields, so that an electronic provenance record could link together all the pieces that led to a result: the data citation (Rule 2), the pointer to the code (Rule 6), the workflow (this rule), and a scholarly paper (Rule 5). But what can you do now? **At a minimum, provide, alongside any deposit of data, a simple sketch of data flow across software, indicating how intermediate and final data products and results are generated. If it's feasible and you are willing to deal with a higher level of complexity, also consider using an online service to encapsulate your workflow (see Links to Useful Resources [C] for a list of services)**. Keep in mind that even if the data used are not "new," in that they come from a well-documented archive, it is still important to document the archive query that produced the data you used, along with all the operations you performed on the data after they were retrieved. Keeping better track of workflow, as context, will likely benefit you and your collaborators enough to justify the loftier, more altruistic, goals espoused here.

## Rule 5. Link Your Data to Your Publications as Often as Possible

Whether your "data" include tables, spreadsheets, images, graphs, databases, and/or code, you should make as much of it as possible available *with* any paper that presents it. **If it's practical and helpful, share your data as early as possible in your research workflow: as soon as you are done with the analysis, even before you write any articles about it.** Your data can even be cited before (or without) its inclusion in a paper (see Rule 7). Many journals now offer standard ways to contribute data to their archives and link it to your paper, often with a persistent identifier. Whenever possible, **embed citations (links) to your data and code, each with its own persistent identifier, right into the text of your paper, just like you would reference other literature.** If a journal hosting your paper doesn't offer a place for your data, and or an identifier for it, use a repository (Rule 8) and get your own identifier (Rule 2). At a *minimum*, you can post, and refer to, a package of files (data, codes, documentation on parameters, metadata, license information, and/or lists of links to such) with a persistent online identifier (Rule 2). And, if your domain's journals' policies do not allow for good data–literature interlinking, try to effect change (see Rules 1 and 10).

## Rule 6. Publish Your Code (Even the Small Bits)

Did you write any code to run your analysis? **No matter how buggy and insignificant you may find it, publish it.** Many easy-to-use source code repositories exist, which allow not only hosting of software but also facilitate collaboration and version tracking (see Links to Useful Resources [D]). Your code, even the shortest script (whether or not you are proud of its quality), can be an important component for understanding your data and how you got your results [6]. Software plays several roles in relation to data and scientific research, and norms around its publication are still evolving and differ across disciplines [7]. In some cases, software is the primary data product (e.g., new algorithms). In some other cases, data are the primary research products, yet the best way to document their provenance is to publish the software that was used to generate them as "metadata." In both cases, publishing the source code and its version history is crucial to enhance transparency and reproducibility. The use of open-source software when possible reduces barriers for subsequent users of your software-related data products [8]. The same best practices discussed above in relation to data and workflow also apply to software materials: cite the software that you use and provide unique, persistent identifiers (Rule 2) to the code you share.

## Rule 7. State How You Want to Get Credit

Chances are that you want to get credit for what you share. The attribution system used for scholarly articles, accomplished via citations, often breaks in the case of data and software. When other authors reuse or cite your data or code, you may get an acknowledgment or an incoming link. If you and your colleagues have gone to the trouble to write a "data paper," whose main purpose is to describe your data and/or code, you may also get a citation [9]. But, "data paper" writing is not always desirable, or relevant. So, how do you go about getting the full credit you deserve for your data and code? **The best way is to simply describe your expectations on how you would like to be acknowledged.** If you want, **you can also release your data under a license and indicate explicitly in the paper or in the metadata how you want others to give you credit.** But, while legal mechanisms have advantages, they can also inadvertently lead to limitations on the reuse of the data you are sharing. In any case, **make information about you (e.g., your name, institution), about the data and/or code (e.g., origin, version, associated files, and metadata), and about exactly how you would like to get credit, as clear as possible.** Easy-to-implement licenses, many of which offer the advantage of being machine-readable, are offered by the Creative Commons organization (http://creativecommons.org/choose/), as are other similar options, such as those offered by Open Data Commons (http://opendatacommons.org/licenses/pddl/). The Links to Useful Resources (G) provides more information.

## Rule 8. Foster and Use Data Repositories

Sometimes the hardest and most time-consuming step of sharing data and code is finding and deciding where to put them. Data-sharing practices vary widely across disciplines: in some fields data sharing and reuse are essential and commonplace, while in others data sharing is a "gift exchange" culture [10]. **If your community already has a standard repository, use it. If you don't know where to start looking, or you need help choosing among relevant repositories, ask an information specialist, such as a data scientist or a librarian working in your field** (and consult the directories of data repositories listed in the Links to Useful Resources [B]). When choosing among repositories, try to find the one offering the best combination of ease-of-deposit, community uptake, accessibility, discoverability, value-added curation, preservation infrastructure, organizational persistence, and support for the data formats and standards you use. **Remember that even if your field has no domain-based repository, your institution may have one**, and your local librarian or archivist can instruct you on how to use that local resource. If neither your community nor your institution has a relevant repository, try a generic repository or consider setting up your own (see Rule 2, and Links to Useful Resources [F]).

## Rule 9. Reward Colleagues Who Share Their Data Properly

Whether you do it in person at scientific meetings and conferences or by written communication when reviewing papers and grants, **reward your colleagues who share data and code. Rally your colleagues and engage your community by providing feedback on the quality of the data assets in your field. Praise those following the best practices.** The more the data created by your colleagues is accessible as an organized collection of some sort, the better your community's research capacity. The more data get shared, used, and cited, the more they improve. Besides personal involvement and encouragement, the best way to reward data sharing is by attribution: always cite the sources of data that you use. **Follow good scientific practice and give credit to those whose data you use, following their preferred reference format and according to current best practices.** Standards and practices for citing and attributing data sources are actively being developed through international partnerships [11,12].

## Rule 10. Be a Booster for Data Science

As Rule 1 says, it is important not just that *you* love your own data, but that *others* love data, too. An attitude that data and code are "second-class objects," behind traditional scholarly publications, is still prevalent. But, every day, as scientists try to use the frustrating but tantalizing hodgepodge of research data available via the present ad hoc network of online systems, the value of organizing an open network of reusable data and code is becoming more and more clear, to more and more people. **You, as a scientist, need to help organize your discipline and your institution to move more quickly toward a world of open, discoverable, reproducible data and research. One important step is to *advocate* for hiring data specialists and for the overall support of institutional programs that improve data sharing.** Make sure not only advanced researchers (e.g., postdocs) experience the pleasures of doing research with freely available data and tools: ***explain* and *show* the value of well-loved data to graduate and undergraduate researchers**. *Teach* whole courses, or mini-courses, related to caring for data and software, or incorporate the ideas into existing courses. *Form groups* specific to your discipline to foster data and code sharing. Hold birds-of-a-feather or special *sessions during large meetings* demonstrating examples in which good sharing practices have led to better results and collaborations. Lead by practicing what you preach.

## Links to Useful Resources

### A: General Data Repositories

- Dataverse (http://thedata.org): A repository for research data that takes care of long-term preservation and good archival practices, while researchers can share, keep control of, and get recognition for their data.
- FigShare (http://figshare.com): A repository where users can make all of their research outputs available in a citable, shareable, and discoverable manner.
- Zenodo (http://zenodo.org): A repository service that enables researchers, scientists, projects, and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of existing institutional or subject-based repositories.
- Dryad (http://datadryad.org): A repository that aims to make data archiving as simple and as rewarding as possible through a suite of services not necessarily provided by publishers or institutional websites.

### B: Directories of Research Data Repositories

- DataBib (http://databib.org): Databib is a tool for helping people identify and locate online repositories of research data. Users and bibliographers create and curate records that describe data repositories that users can search.
- Re3data.org (http://www.re3data.org): Re3data is a global registry of research data repositories from different academic disciplines for researchers, funding bodies, publishers, and scholarly institutions.
- Open Access Directory (http://oad. simmons.edu/oadwiki/Data_repositories): A list of repositories and databases for open data.
- Force 11 Catalog (http://www.force11.org/catalog): A dynamic inventory of web-based scholarly resources, a collection of alternative

publication systems, databases, organizations and groups, software, services, standards, formats, and training tools.

### C: Workflow Management Systems

- Taverna (http://www.taverna.org.uk): An open-source and domain-independent workflow management system—a suite of tools used to design and execute scientific workflows and aid in silico experimentation.
- Kepler (https://kepler-project.org): Software designed to help scientists, analysts, and computer programmers create, execute, and share models and analyses across a broad range of scientific and engineering disciplines.
- Wings (http://www.wings-workflows.org): A semantic workflow system that assists scientists with the design of computational experiments.
- VisTrails (http://www.vistrails.org): An open-source scientific workflow and provenance management system that supports data exploration and visualization.
- Knime (http://www.knime.org): A graphical workbench for the entire analysis process: data access, data transformation, initial investigation, powerful predictive analytics, visualization, and reporting.

### D: Source Code Repositories

- Github (http://github.com): A web-based hosting service for software development projects that use the Git revision control system, including many open-source projects.
- Git (http://git-scm.com): A free and open-source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.
- Mercurial (http://mercurial.selenic.com): A free, distributed source control management tool. It efficiently handles projects of any size and offers an easy and intuitive interface.
- BitBucket (https://bitbucket.org): A web-based hosting service for projects that use either the Mercurial or Git revision control systems.

### E: Systems to Package, Access, and Execute Data and Code

- IPython Notebook (http://ipython.org/notebook.html): A web-based interactive computational environment where you can combine code execution, text, mathematics, plots, and rich media into a single document.
- ROpenSci (http://ropensci.org): A suite of packages that allow access to data repositories through the R statistical programming environment.
- Authorea (https://authorea.com): A collaborative online word processor for scholarly papers that allows the writing of web-native, living, dynamic, "executable" articles that include text, mathematical notation, images, and data. It currently supports inclusion and rendering of d3.js plots and IPython notebooks.
- Dexy (http://dexy.it): A multipurpose project automation tool for working with documents via a command-line interface.

### F: Software Tools to Run Your Own Document Repository

- Invenio (http://invenio-software.org): Invenio is a free software suite enabling you to run your own digital library or document repository on the web. Invenio is an ideal solution for running document repositories of moderate to large sizes (several millions of records). Invenio is codeveloped by CERN, DESY, EPFL, FNAL, and SLAC.
- Eprints (http://www.eprints.org/software): EPrints is one of the easiest and fastest ways to set up small to medium-sized repositories of open-access research literature, scientific data, theses, reports, and multimedia. Developed at the University of Southampton, UK.
- DSpace (http://www.dspace.org): DSpace is a turnkey institutional repository application developed by the Duraspace organization.

### G: Licensing and Privacy

- Open Source Initiative (http://opensource.org/licenses): Open-source licenses are licenses that comply with the Open Source Definition: they allow software to be freely used, modified, and shared. These include Apache, BSD, GNU (GPL), MIT, and the Mozilla Public License.
- Privacy Tools for Sharing Research Data (http://privacytools.seas.harvard.edu): A Harvard-based collaborative and multidisciplinary effort to help enable the collection, analysis, and sharing of personal data for research in social science and other fields while providing privacy for individual subjects.


## Acknowledgments

This article is an outcome of an Exploratory Seminar called "What to Keep and How to Analyze It: Data Curation and Data Analysis with Multiple Phases" (http://www.radcliffe.harvard.edu/exploratory-seminars/what-to-keep-and-how-to-analyze-it) organized by Xiao-Li Meng and Alyssa Goodman, held on May 9–10, 2013 at the **Radcliffe Institute for Advanced Study, Harvard University**, Cambridge, Massachusetts. This article was written **collaboratively**, **online**, in the **open**, using Authorea. Every iteration of the writing procedure is logged and available in the online version of this article at https://www.authorea.com/3410.


## References


1. Galilei G (1610) Sidereus nuncius. Mapping of the stars series. Thomas Baglioni.
2. Galilei G (1618) The Assayer, as translated by Stillman Drake (1957). Available: http://books.google.com/books?id=uSctMwEACAAJ.
3. Drake S (1957) Discoveries and opinions of Galileo: including The Starry Messenger (1610), Letter to the Grand Duchess Christina (1615), and Excerpts from Letters on Sunspots (1613), The Assayer (1623). Anchor Books. Available: http://books.google.com/books?id=P2tmjengyBcC.
4. Holdren J (2013) Increasing public access to the results of scientific research. Memorandum of the US Office of Science and Technology, 22 February 2013.
5. Wren JD (2008) URL decay in MEDLINE–a 4-year follow-up study. Bioinformatics 24: 1381–1385.
6. Barnes N (2010) Publish your computer code: it is good enough. Nature 467: 753–753.
7. Shamir L, Wallin JF, Allen A, Berriman B, Teuben P, et al. (2013) Practices in source code sharing in astrophysics. Astronomy and Computing 1: 54–58.
8. Prlić A, Procter JB (2012) Ten simple rules for the open development of scientific software. PLoS Comput Biol 8: e1002802.
9. Chavan V, Penev L (2011) The data paper: a mechanism to incentivize data publishing in biodi-versity science. BMC Bioinformatics 12: S2.
10. Wallis JC, Rolando E, Borgman CL (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS ONE 8: e67332.
11. Uhlir PE (2012) For attribution – developing data attribution and citation practices and standards: summary of an international workshop. The National Academies Press.
12. FORCE11 (2013) Amsterdam Manifesto on Data Citation Principles. Available: http://www.force11.org/AmsterdamManifesto.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Effective Computational Research

**James M. Osborne[1,2]*, Miguel O. Bernabeu[3,4], Maria Bruna[1,2], Ben Calderhead[3], Jonathan Cooper[1], Neil Dalchau[2], Sara-Jane Dunn[2], Alexander G. Fletcher[5], Robin Freeman[2,3], Derek Groen[4], Bernhard Knapp[6], Greg J. McInerny[1,2], Gary R. Mirams[1], Joe Pitt-Francis[1], Biswa Sengupta[7], David W. Wright[3,4], Christian A. Yates[5], David J. Gavaghan[1], Stephen Emmott[2], Charlotte Deane[6]**

**1** Computational Biology Group, Department of Computer Science, University of Oxford, Wolfson Building, Oxford, United Kingdom, **2** Computational Science Laboratory, Microsoft Research, Cambridge, United Kingdom, **3** CoMPLEX, Mathematical and Physical Sciences, University College London, Physics Building, London, United Kingdom, **4** Centre for Computational Science, Department of Chemistry, University College London, London, United Kingdom, **5** Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Oxford, United Kingdom, **6** Department of Statistics, University of Oxford, Oxford, United Kingdom, **7** The Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, United Kingdom

In order to attempt to understand the complexity inherent in nature, mathematical, statistical and computational techniques are increasingly being employed in the life sciences. In particular, the use and development of software tools is becoming vital for investigating scientific hypotheses, and a wide range of scientists are finding software development playing a more central role in their day-to-day research. In fields such as biology and ecology, there has been a noticeable trend towards the use of quantitative methods for both making sense of ever-increasing amounts of data [1] and building or selecting models [2].

As Research Fellows of the "2020 Science" project (http://www.2020science.net), funded jointly by the EPSRC (Engineering and Physical Sciences Research Council) and Microsoft Research, we have firsthand experience of the challenges associated with carrying out multidisciplinary computation-based science [3–5]. In this paper we offer a jargon-free guide to best practice when developing and using software for scientific research. While many guides to software development exist, they are often aimed at computer scientists [6] or concentrate on large open-source projects [7]; the present guide is aimed specifically at the vast majority of scientific researchers: those without formal training in computer science. We present our ten simple rules with the aim of enabling scientists to be more effective in undertaking research and therefore maximise the impact of this research within the scientific community. While these rules are described individually, collectively they form a single vision for how to approach the practical side of computational science.

Our rules are presented in roughly the chronological order in which they should be undertaken, beginning with things that, as a computational scientist, you should do *before* you even think about writing any code. For each rule, guides on getting started, links to relevant tutorials, and further reading are provided in the supplementary material (Text S1).

## Rule 1: Look Before You Leap

One of the key considerations in the development of any method, computational or otherwise, is whether it has previously been approached by someone else. A growing wealth of software toolboxes and libraries exist to tackle many problems. However, assessing the range and quality of what is available can be hard, especially when addressing nontraditional problems. A simple but often-overlooked approach is to conduct a software literature review to ascertain what software is available and has been successfully employed. Software repositories (e.g., GitHub, https://github.com/, and SourceForge, http://sourceforge.net/) are a good place to begin a review. Furthermore, engaging with the network of researchers surrounding your own is invaluable; see [8] and [9] for advice on this. If your coworkers write software in the same language or use particular toolboxes, you may be able to consult their expertise in order to accelerate and provide support for your work.

## Rule 2: Develop a Prototype First

Before writing any code, it is imperative to clarify what you are trying to implement: what functionality do you require, and what interfaces do you need? When implementing your latest developments, you should first begin by considering a prototype (i.e., a simplified version of the full system or algorithm) to gain insight and to guide the next steps. This is equally relevant whether building on existing code or starting from scratch. By prototyping new functionality and building code up incrementally, you can check that each element of your code operates as expected (and each incremental development can be tested; see Rule 8). Breaking your problem up into smaller elements like this will also help to provide structure to your code and will make it much easier when you subsequently need to extend it. From a practical point of view, it will usually be easier to prototype mathematical and statistical methods in a "higher-level" language, for example Matlab, R, or Python. Although these languages can be slower to run than optimized code in a

**Citation:** Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, et al. (2014) Ten Simple Rules for Effective Computational Research. PLoS Comput Biol 10(3): e1003506. doi:10.1371/journal.pcbi.1003506

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** March 27, 2014



**Funding:** This work is part of the 2020 Science programme, which is funded through the EPSRC Cross-Disciplinary Interface Programme (grant number EP/I017909/1) and is also supported by Microsoft Research Ltd. The funders had no role in the preparation of the manuscript.

**Competing Interests:** ND, SJD, and SE are paid employees of Microsoft Research (MSR) and JMO, MB, RF, and GJM are partly supported by MSR. This does not alter our adherence to the PLOS policies on sharing data and materials. All other authors have declared that no competing interests exist.

* E-mail: James.Osborne@cs.ox.ac.uk

"lower-level" language, their straightforward nature, built-in functionality, and available libraries mean that you will spend less time expressing your ideas in code and searching for bugs.

## Rule 3: Make Your Code Understandable to Others (and Yourself)

When revising or adapting existing code, the absence of documentation and comments can result in errors and time drains. Such documentation not only makes your code more understandable to others but also to your future self (put simply, the code tells you "how", the comments tell you "why"). The program code itself can be made more understandable by using meaningful variable names and formatting the code consistently. While commenting and documentation is often neglected when faced with deadlines, developing and maintaining a standardised way of commenting your code will be of great benefit. As well as low-level documentation in the code, you should maintain a record of the "big picture" functionality (i.e., interconnectivity of components and input/output formats). This could take the form of a high-level diagram or description of the system, whether by hand on paper, in verbose code comments, or using standardized approaches such as UML (Unified Modelling Language) (see Text S1). When you are reviewing your code for documentation you should actively seek ways to break it up into modules. This not only aids structure and readability but also avoids the error-prone and tedious task of debugging and updating two (or more) copies of the same code. As a rule of thumb, if you write the same code twice, it should become a function, subroutine, or method.

## Rule 4: Don't Underestimate the Complexity of Your Task

When developing your code, you should keep a record of your work. This could be in the form of a "logbook" file or a paper notebook where you store commonly used commands and other notes; another good option is an online tool such as Evernote (http://evernote.com/). You will often find that you have to choose between spending a long time doing a task by hand and possibly spending longer learning how to automate it. In order to automate the task, you will probably need to learn how to use some basic tools such as text editors or scripting languages. Don't be tempted to think, "This is just a one-off, I'll get on with it;" it won't be. You will find bugs, wish to change a parameter, or need to alter a figure slightly, and you will eventually have to repeat the whole process. Even if you are *certain* that it really is a one-off task, use your "logbook" and keep a record of the list of commands you used, since this is the first step towards automating the task if and when the time comes. However, it is not appropriate to automate everything, and you need to find a good balance, automating opportunistically, taking the expected time and cost into account. A good rule to follow is "the rule of three:" once you have had to do the same thing twice already, automate it.

## Rule 5: Understand the Mathematical, Numerical, and Computational Methods Underpinning Your Work

When solving any computational model, you should always ensure that you are using the appropriate numerical method for your problem, and that any constraints and conditions are satisfied. A basic understanding of numerical analysis and, in particular, the concepts of rate of convergence, order, and stability of numerical methods will pay dividends. Care should also be taken to ensure that any assumptions made in the derivation of the underlying mathematical models or methods (e.g., having a sufficiently large number of objects to permit a continuum approximation) hold for all system states of interest. You should consult the relevant literature (and communities) that explains these methods and their advantages and/or disadvantages and not steam ahead without first gaining an understanding of which methods are appropriate. By fully understanding the mathematical and numerical methods being used, you can be confident that your results reflect the true behaviour of the underlying model and are not numerical or computational artefacts.

## Rule 6: Use Pictures: They Really Are Worth a Thousand Words

Visualisation and graphics are fundamental to developing, understanding, and testing hypotheses, and are indispensable for verifying and validating computational methods (e.g., revealing correlations, covariation, position, structure, flows, orientation, anomalies, and outliers). So, from day one, spend time developing the visual components of your work. Learn, develop and use visualisation software and tools to ensure that you understand your research outputs and can effectively communicate your findings. You may well need to develop novel visualisations for your work, but keep the basic figures. You needed them to understand your results, model, and implementation, and so will anyone else. You should ensure that your visualisation algorithms can be executed separately so that they can be reused by you and others (for the same and different tasks) and refined for other formats (e.g., publications, presentations, and websites). In reality, all scientists could be better educated in design, so any investment will be rewarded, especially by receiving feedback on visualization from users.

## Rule 7: Version Control *Everything*

Version control systems (VCSs) offer an easy way to store and back up not only the current version of your code that you are working on but also every previous version of the code (in what's known as a repository). This not only saves you from having to keep multiple copies of the same file but also allows you to "roll back" to an older "working" version of the code if things go wrong. VCSs also allow you to share material between multiple machines, operating systems, and more importantly, users in a simple and robust manner. Two of the most popular VCSs are Subversion (http://subversion.apache.org) and Git (http://www.github.com), both of which offer many advanced features for managing your code. Cloud storage such as Dropbox (http://www.dropbox.com) and SkyDrive (http://www.skydrive.live.com) offer basic file sharing and backup facilities; however, they don't offer the code management features of true VCSs, so the effort put in to learning a VCS is well worth it (see Text S1 for guides on getting started with VCSs). While the primary use of version control is to manage the development and distribution of code, many other collaborative endeavours can be stored in a version control repository. In particular, using version control tools while preparing publications can save time and effort, especially when dealing with input from multiple authors. For example, contributions to this manuscript were managed using a VCS.

## Rule 8: Test *Everything*

Any non-trivial computer program will have bugs when first written, often subtle ones that are hard to detect, which may lead to incorrect results. Indeed, in extreme cases this has caused high-profile retractions of papers [10]. Simple tests that the software behaviour matches expectations are essential for ensuring robust results, minimising the presence of bugs, and gaining confidence in your code (for you and others). As a result of the time pressures inherent in academia, often software testing is performed manually in an *ad hoc* manner, to determine whether results ''look roughly right'' [11]. However, a systematic approach to testing pays dividends. You should learn how to test effectively to avoid the illusion of reliability. For example, compare low-level routines against analytical or prototype solutions (see Rule 2) or experimental data and consider ''corner cases'' and both branches of ''if'' statements. Get the computer to run tests for you automatically and alert you to problems, using a suitable testing framework (see Text S1). Ideally this should be tied to a version control system (see Rule 7) so that tests are run automatically whenever new code is committed to the repository. A useful rule is to turn bugs you fix into new tests to avoid them recurring. Testing gives you the confidence to modify your code without worrying that you are breaking it. Testing can also provide a means for reproducing results of published papers. By setting up a test comparing against published values, you can easily find out when fixing a newly identified bug changes published results.

## Rule 9: Share *Everything*

Just as it is a common practice to publish your research findings in peer-reviewed journals, if an important part of your research involves developing new software tools and/or collecting new data, you should consider sharing these [7]. Based on our collective experience, we advocate an open approach of sharing source code, data, and results as freely as possible. You should ask yourself, ''Why not share?'' If the answer is, ''I am worried that people would find mistakes in it,'' then, as a scientist, this should be the strongest argument in favour of sharing it! The provision of such resources openly provides the means to replicate, reproduce, and examine newly developed methods and techniques. Open sharing not only facilitates the scientific enterprise through replication, validation, and error checking, but also deters fraud and malpractice through transparency. It is our opinion that the many arguments in favour of openly sharing code, data, and results far outweigh any against. In many modern computational analyses, the source code represents a readable, executable methodology of the research in question. Sharing is the key to a sustainable future for computational science, and publishers are beginning to require it, with some considering reviewing the software used to generate results [12].

## Rule 10: Keep Going!

Our advice arises from our collective experience, and we continue to strive to obey these rules in our work. Scientists have a wide variety of demands on their time (researching, writing papers [13], teaching [14], applying for grants, administration, etc.) and have to make the most of limited resources. Becoming more technically effective can seem daunting without strategies for making progress and keeping motivated. So, prioritise in a way that suits you and your projects and career aspirations. One strategy is to implement another of these rules each time you start a new project, to build a growing repertoire rather than trying to do everything at once. Take every opportunity to teach and help others to do what you have learnt.

## Supporting Information

**Text S1** Supplementary material for paper. Includes guides for getting started with each rule, along with references to useful links and further reading.
(PDF)

## References


1. Kumar S, Dudley J (2007) Bioinformatics software for biologists in the genomics era. Bioinformatics 23: 1713–1717. doi:10.1093/bioinformatics/btm239
2. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, et al. (2012) A Whole-Cell Computational Model Predicts Phenotype from Genotype. Cell 150: 389–401. doi:10.1016/j.cell.2012.05.044
3. Mirams GR, Arthurs CJ, Bernabeu MO, Bordas R, Cooper J, et al. (2013) Chaste: an open source C++ library for computational physiology and biology. PLOS Comput Biol 9: e1002970. doi:10.1371/journal.pcbi.1002970
4. Dalchau N, Phillips A, Goldstein LD, Howarth M, Cardelli L, et al. (2011) A peptide filtering relation quantifies MHC class I peptide optimization. PLOS Comput Biol 7: e1002144. doi:10.1371/journal.pcbi.1002144
5. Bernabeu MO, Nash RW, Groen D, Carver HB, Hetherington J, et al. (2013) Impact of blood rheology on wall shear stress in a model of the middle cerebral artery. Interface Focus 3: 20120094. doi:10.1098/rsfs.2012.0094
6. Mozilla Science Lab (2013) Software Carpentry. Available: http://software-carpentry.org/. Accessed 18 March 2013.
7. Prlić A, Procter JB (2012) Ten simple rules for the open development of scientific software. PLOS Comput Biol 8: e1002802. doi:10.1371/journal.pcbi.1002802
8. Dall'Olio GM, Marino J, Schubert M, Keys KL, Stefan MI, et al. (2011) Ten simple rules for getting help from online scientific communities. PLOS Comput Biol 7: e1002202. doi:10.1371/journal.pcbi.1002202
9. Michaut M (2011) Ten simple rules for getting involved in your scientific community. PLOS Comput Biol 7: e1002232. doi:10.1371/journal.pcbi.1002232
10. Chang G, Roth CB, Reyes CL, Pornillos O, Chen Y, et al. (2006) Retraction. Science 314: 1875. doi:10.1126/science.314.5807.1875b
11. Pitt-Francis J, Bernabeu MO, Cooper J, Garny A, Momtahan L, et al. (2008) Chaste: Using agile programming techniques to develop computational biology software. Phil Trans R Soc A 366: 3111–3136. doi:10.1098/rsta.2008.0096
12. Hayden EC (2013) Mozilla plan seeks to debug scientific code. Nature 501: 472. doi:10.1038/501472a
13. Bourne PE (2005) Ten simple rules for getting published. PLOS Comput Biol 1: e57. doi:10.1371/journal.pcbi.0010057
14. Vicens Q, Bourne PE (2009) Ten simple rules to combine teaching and research. PLOS Comput Biol 5: e1000358. doi:10.1371/journal.pcbi.1000358

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Reproducible Computational Research

**Geir Kjetil Sandve[1,2]*, Anton Nekrutenko[3], James Taylor[4], Eivind Hovig[1,5,6]**

**1** Department of Informatics, University of Oslo, Blindern, Oslo, Norway, **2** Centre for Cancer Biomedicine, University of Oslo, Blindern, Oslo, Norway, **3** Department of Biochemistry and Molecular Biology and The Huck Institutes for the Life Sciences, Penn State University, University Park, Pennsylvania, United States of America, **4** Department of Biology and Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia, United States of America, **5** Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Montebello, Oslo, Norway, **6** Institute for Medical Informatics, The Norwegian Radium Hospital, Oslo University Hospital, Montebello, Oslo, Norway

Replication is the cornerstone of a cumulative science [1]. However, new tools and technologies, massive amounts of data, interdisciplinary approaches, and the complexity of the questions being asked are complicating replication efforts, as are increased pressures on scientists to advance their research [2]. As full replication of studies on independently collected data is often not feasible, there has recently been a call for reproducible research as an attainable minimum standard for assessing the value of scientific claims [3]. This requires that papers in experimental science describe the results and provide a sufficiently clear protocol to allow successful repetition and extension of analyses based on original data [4].

The importance of replication and reproducibility has recently been exemplified through studies showing that scientific papers commonly leave out experimental details essential for reproduction [5], studies showing difficulties with replicating published experimental results [6], an increase in retracted papers [7], and through a high number of failing clinical trials [8,9]. This has led to discussions on how individual researchers, institutions, funding bodies, and journals can establish routines that increase transparency and reproducibility. In order to foster such aspects, it has been suggested that the scientific community needs to develop a "culture of reproducibility" for computational science, and to require it for published claims [3].

We want to emphasize that reproducibility is not only a moral responsibility with respect to the scientific field, but that a lack of reproducibility can also be a burden for you as an individual researcher. As an example, a good practice of reproducibility is necessary in order to allow previously developed methodology to be effectively applied on new data, or to allow reuse of code and results for new projects. In other words, good habits of reproducibility may actually turn out to be a time-saver in the longer run.

We further note that reproducibility is just as much about the habits that ensure reproducible research as the technologies that can make these processes efficient and realistic. Each of the following ten rules captures a specific aspect of reproducibility, and discusses what is needed in terms of information handling and tracking of procedures. If you are taking a bare-bones approach to bioinformatics analysis, i.e., running various custom scripts from the command line, you will probably need to handle each rule explicitly. If you are instead performing your analyses through an integrated framework (such as GenePattern [10], Galaxy [11], LONI pipeline [12], or Taverna [13]), the system may already provide full or partial support for most of the rules. What is needed on your part is then merely the knowledge of how to exploit these existing possibilities.

In a pragmatic setting, with publication pressure and deadlines, one may face the need to make a trade-off between the ideals of reproducibility and the need to get the research out while it is still relevant. This trade-off becomes more important when considering that a large part of the analyses being tried out never end up yielding any results. However, frequently one will, with the wisdom of hindsight, contemplate the missed opportunity to ensure reproducibility, as it may already be too late to take the necessary notes from memory (or at least much more difficult than to do it while underway). We believe that the rewards of reproducibility will compensate for the risk of having spent valuable time developing an annotated catalog of analyses that turned out as blind alleys.

As a minimal requirement, you should at least be able to reproduce the results yourself. This would satisfy the most basic requirements of sound research, allowing any substantial future questioning of the research to be met with a precise explanation. Although it may sound like a very weak requirement, even this level of reproducibility will often require a certain level of care in order to be met. There will for a given analysis be an exponential number of possible combinations of software versions, parameter values, preprocessing steps, and so on, meaning that a failure to take notes may make exact reproduction essentially impossible.

With this basic level of reproducibility in place, there is much more that can be wished for. An obvious extension is to go from a level where you can reproduce results in case of a critical situation to a level where you can practically and routinely reuse your previous work and increase your productivity. A second extension is to ensure that peers have a practical possibility of reproducing your results, which can lead to increased trust in, interest for, and citations of your work [6,14].

**Citation:** Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol 9(10): e1003285. doi:10.1371/journal.pcbi.1003285

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** October 24, 2013



**Funding:** The authors' laboratories are supported by US National Institutes of Health grants HG005133, HG004909, and HG006620 and US National Science Foundation grant DBI 0850103. Additional funding is provided, in part, by the Huck Institutes for the Life Sciences at Penn State, the Institute for Cyberscience at Penn State, and a grant with the Pennsylvania Department of Health using Tobacco Settlement Funds. The funders had no role in the preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: geirksa@ifi.uio.no

We here present ten simple rules for reproducibility of computational research. These rules can be at your disposal for whenever you want to make your research more accessible—be it for peers or for your future self.

## Rule 1: For Every Result, Keep Track of How It Was Produced

Whenever a result may be of potential interest, keep track of how it was produced. When doing this, one will frequently find that getting from raw data to the final result involves many interrelated steps (single commands, scripts, programs). We refer to such a sequence of steps, whether it is automated or performed manually, as an analysis workflow. While the essential part of an analysis is often represented by only one of the steps, the full sequence of pre- and post-processing steps are often critical in order to reach the achieved result. For every involved step, you should ensure that every detail that may influence the execution of the step is recorded. If the step is performed by a computer program, the critical details include the name and version of the program, as well as the exact parameters and inputs that were used.

Although manually noting the precise sequence of steps taken allows for an analysis to be reproduced, the documentation can easily get out of sync with how the analysis was really performed in its final version. By instead specifying the full analysis workflow in a form that allows for direct execution, one can ensure that the specification matches the analysis that was (subsequently) performed, and that the analysis can be reproduced by yourself or others in an automated way. Such executable descriptions [10] might come in the form of simple shell scripts or makefiles [15,16] at the command line, or in the form of stored workflows in a workflow management system [10,11,13,17,18].

As a minimum, you should at least record sufficient details on programs, parameters, and manual procedures to allow yourself, in a year or so, to approximately reproduce the results.

## Rule 2: Avoid Manual Data Manipulation Steps

Whenever possible, rely on the execution of programs instead of manual procedures to modify data. Such manual procedures are not only inefficient and error-prone, they are also difficult to reproduce. If working at the UNIX command line, manual modification of files can usually be replaced by the use of standard UNIX commands or small custom scripts. If working with integrated frameworks, there will typically be a quite rich collection of components for data manipulation. As an example, manual tweaking of data files to attain format compatibility should be replaced by format converters that can be reenacted and included into executable workflows. Other manual operations like the use of copy and paste between documents should also be avoided. If manual operations cannot be avoided, you should as a minimum note down which data files were modified or moved, and for what purpose.

## Rule 3: Archive the Exact Versions of All External Programs Used

In order to exactly reproduce a given result, it may be necessary to use programs in the exact versions used originally. Also, as both input and output formats may change between versions, a newer version of a program may not even run without modifying its inputs. Even having noted which version was used of a given program, it is not always trivial to get hold of a program in anything but the current version. Archiving the exact versions of programs actually used may thus save a lot of hassle at later stages. In some cases, all that is needed is to store a single executable or source code file. In other cases, a given program may again have specific requirements to other installed programs/packages, or dependencies to specific operating system components. To ensure future availability, the only viable solution may then be to store a full virtual machine image of the operating system and program. As a minimum, you should note the exact names and versions of the main programs you use.

## Rule 4: Version Control All Custom Scripts

Even the slightest change to a computer program can have large intended or unintended consequences. When a continually developed piece of code (typically a small script) has been used to generate a certain result, only that exact state of the script may be able to produce that exact output, even given the same input data and parameters. As also discussed for rules 3 and 6, exact reproduction of results may in certain situations be essential. If computer code is not systematically archived along its evolution, backtracking to a code state that gave a certain result may be a hopeless task. This can cast doubt on previous results, as it may be impossible to know if they were partly the result of a bug or otherwise unfortunate behavior.

The standard solution to track evolution of code is to use a version control system [15], such as Subversion, Git, or Mercurial. These systems are relatively easy to set up and use, and may be used to systematically store the state of the code throughout development at any desired time granularity.

As a minimum, you should archive copies of your scripts from time to time, so that you keep a rough record of the various states the code has taken during development.

## Rule 5: Record All Intermediate Results, When Possible in Standardized Formats

In principle, as long as the full process used to produce a given result is tracked, all intermediate data can also be regenerated. In practice, having easily accessible intermediate results may be of great value. Quickly browsing through intermediate results can reveal discrepancies toward what is assumed, and can in this way uncover bugs or faulty interpretations that are not apparent in the final results. Secondly, it more directly reveals consequences of alternative programs and parameter choices at individual steps. Thirdly, when the full process is not readily executable, it allows parts of the process to be rerun. Fourthly, when reproducing results, it allows any experienced inconsistencies to be tracked to the steps where the problems arise. Fifth, it allows critical examination of the full process behind a result, without the need to have all executables operational. When possible, store such intermediate results in standardized formats. As a minimum, archive any intermediate result files that are produced when running an analysis (as long as the required storage space is not prohibitive).

## Rule 6: For Analyses That Include Randomness, Note Underlying Random Seeds

Many analyses and predictions include some element of randomness, meaning the same program will typically give slightly different results every time it is executed (even when receiving identical inputs and parameters). However, given the same initial seed, all random numbers used in an analysis will be equal, thus giving identical results every time it is run. There is a large difference between observing that a result has been reproduced exactly or only approximately. While achieving equal results is a strong indication that a procedure has been reproduced exactly, it is often hard to conclude anything when achieving only approximately equal results. For analyses that involve random numbers, this means that the random seed should be recorded. This allows results to be reproduced exactly by providing the same seed to the random number generator in future runs. As a minimum, you should note which analysis steps involve randomness, so that a certain level of discrepancy can be anticipated when reproducing the results.

## Rule 7: Always Store Raw Data behind Plots

From the time a figure is first generated to it being part of a published article, it is often modified several times. In some cases, such modifications are merely visual adjustments to improve readability, or to ensure visual consistency between figures. If raw data behind figures are stored in a systematic manner, so as to allow raw data for a given figure to be easily retrieved, one can simply modify the plotting procedure, instead of having to redo the whole analysis. An additional advantage of this is that if one really wants to read fine values in a figure, one can consult the raw numbers. In cases where plotting involves more than a direct visualization of underlying numbers, it can be useful to store both the underlying data and the processed values that are directly visualized. An example of this is the plotting of histograms, where both the values before binning (original data) and the counts per bin (heights of visualized bars) could be stored. When plotting is performed using a command-based system like R, it is convenient to also store the code used to make the plot. One can then apply slight modifications to these commands, instead of having to specify the plot from scratch. As a minimum, one should note which data formed the basis of a given plot and how this data could be reconstructed.

## Rule 8: Generate Hierarchical Analysis Output, Allowing Layers of Increasing Detail to Be Inspected

The final results that make it to an article, be it plots or tables, often represent highly summarized data. For instance, each value along a curve may in turn represent averages from an underlying distribution. In order to validate and fully understand the main result, it is often useful to inspect the detailed values underlying the summaries. A common but impractical way of doing this is to incorporate various debug outputs in the source code of scripts and programs. When the storage context allows, it is better to simply incorporate permanent output of all underlying data when a main result is generated, using a systematic naming convention to allow the full data underlying a given summarized value to be easily found. We find hypertext (i.e., html file output) to be particularly useful for this purpose. This allows summarized results to be generated along with links that can be very conveniently followed (by simply clicking) to the full data underlying each summarized value. When working with summarized results, you should as a minimum at least once generate, inspect, and validate the detailed values underlying the summaries.

## Rule 9: Connect Textual Statements to Underlying Results

Throughout a typical research project, a range of different analyses are tried and interpretation of the results made. Although the results of analyses and their corresponding textual interpretations are clearly interconnected at the conceptual level, they tend to live quite separate lives in their representations: results usually live on a data area on a server or personal computer, while interpretations live in text documents in the form of personal notes or emails to collaborators. Such textual interpretations are not generally mere shadows of the results—they often involve viewing the results in light of other theories and results. As such, they carry extra information, while at the same time having their necessary support in a given result.

If you want to reevaluate your previous interpretations, or allow peers to make their own assessment of claims you make in a scientific paper, you will have to connect a given textual statement (interpretation, claim, conclusion) to the precise results underlying the statement. Making this connection when it is needed may be difficult and error-prone, as it may be hard to locate the exact result underlying and supporting the statement from a large pool of different analyses with various versions.

To allow efficient retrieval of details behind textual statements, we suggest that statements are connected to underlying results already from the time the statements are initially formulated (for instance in notes or emails). Such a connection can for instance be a simple file path to detailed results, or the ID of a result in an analysis framework, included within the text itself. For an even tighter integration, there are tools available to help integrate reproducible analyses directly into textual documents, such as Sweave [19], the GenePattern Word add-in [4], and Galaxy Pages [20]. These solutions can also subsequently be used in connection with publications, as discussed in the next rule.

As a minimum, you should provide enough details along with your textual interpretations so as to allow the exact underlying results, or at least some related results, to be tracked down in the future.

## Rule 10: Provide Public Access to Scripts, Runs, and Results

Last, but not least, all input data, scripts, versions, parameters, and intermediate results should be made publicly and easily accessible. Various solutions have now become available to make data sharing more convenient, standardized, and accessible in particular domains, such as for gene expression data [21–23]. Most journals allow articles to be supplemented with online material, and some journals have initiated further efforts for making data and code more integrated with publications [3,24]. As a minimum, you should submit the main data and source code as supplementary material, and be prepared to respond to any requests for further data or methodology details by peers.

Making reproducibility of your work by peers a realistic possibility sends a strong signal of quality, trustworthiness, and transparency. This could increase the quality and speed of the reviewing process on your work, the chances of your work getting published, and the chances of your work being taken further and cited by other researchers after publication [25].

## References


1. Crocker J, Cooper ML (2011) Addressing scientific fraud. Science 334: 1182.
2. Jasny BR, Chin G, Chong L, Vignieri S (2011) Data replication & reproducibility. Again, and again, and again…. Introduction. Science 334: 1225.
3. Peng RD (2011) Reproducible research in computational science. Science 334: 1226–1227.
4. Mesirov JP (2010) Computer science. Accessible reproducible research. Science 327: 415–416.
5. Nekrutenko A, Taylor J (2012) Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet 13: 667–672.
6. Ioannidis JP, Allison DB, Ball CA, Coulibaly I, Cui X, et al. (2009) Repeatability of published microarray gene expression analyses. Nat Genet 41: 149–155.
7. Steen RG (2011) Retractions in the scientific literature: is the incidence of research fraud increasing? J Med Ethics 37: 249–253.
8. Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov 10: 712.
9. Begley CG, Ellis LM (2012) Drug development: raise standards for preclinical cancer research. Nature 483: 531–533.
10. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, et al. (2006) GenePattern 2.0. Nat Genet 38: 500–501.
11. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15: 1451–1455.
12. Rex DE, Ma JQ, Toga AW (2003) The LONI Pipeline Processing Environment. Neuroimage 19: 1033–1048.
13. Oinn T, Addis M, Ferris J, Marvin D, Senger M, et al. (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics 20: 3045–3054.
14. Piwowar HA, Day RS, Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. PLoS ONE 2: e308. doi:10.1371/journal.pone.0000308.
15. Heroux MA, Willenbring JM (2009) Barely sufficient software engineering: 10 practices to improve your cse software. In: 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering. pp. 15–21.
16. Schwab M, Karrenbach M, Claerbout J (2000) Making scientific computations reproducible. Comput Sci Eng 2: 61–67.
17. Goble CA, Bhagat J, Aleksejevs S, Cruickshank D, Michaelides D, et al. (2010) myExperiment: a repository and social network for the sharing of bioinformatics workflows. Nucleic Acids Res 38: W677–682.
18. Deelman E, Singh G, Su M-H, Blythe J, Gil Y, et al. (2005) Pegasus: a framework for mapping complex scientific workflows onto distributed systems. Scientific Programming Journal 13: 219–237.
19. Leisch F (2002) Sweave: dynamic generation of statistical reports using literate data analysis. In: Härdle W, Rönz B, editors. Compstat: proceedings in computational statistics. Heidelberg, Germany: Physika Verlag. pp. 575–580.
20. Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11: R86.
21. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29: 365–371.
22. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, et al. (2003) ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 31: 68–71.
23. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30: 207–210.
24. Sneddon TP, Li P, Edmunds SC (2012) GigaDB: announcing the GigaScience database. Gigascience 1: 11.
25. Prlić A, Procter JB (2012) Ten simple rules for the open development of scientific software. PLoS Comput Biol 8: e1002802. doi:10.1371/journal.pcbi.1002802.

Editorial

# Ten Simple Rules for Doing Your Best Research, According to Hamming

Thomas C. Erren*, Paul Cullen, Michael Erren, Philip E. Bourne

This editorial can be considered the preface to the "Ten Simple Rules" series [1–7]. The rules presented here are somewhat philosophical and behavioural rather than concrete suggestions for how to tackle a particular scientific professional activity such as writing a paper or a grant. The thoughts presented are not our own; rather, we condense and annotate some excellent and timeless suggestions made by the mathematician Richard Hamming two decades ago on how to do "first-class research" [8]. As far as we know, the transcript of the Bell Communications Research Colloquium Seminar provided by Dr. Kaiser [8] was never formally published, so that Dr. Hamming's thoughts are not as widely known as they deserve to be. By distilling these thoughts into something that can be thought of as "Ten Simple Rules," we hope to bring these ideas to broader attention.

Hamming's 1986 talk was remarkable. In "You and Your Research," he addressed the question: How can scientists do great research, i.e., Nobel-Prize-type work? His insights were based on more than forty years of research as a pioneer of computer science and telecommunications who had the privilege of interacting with such luminaries as the physicists Richard Feynman, Enrico Fermi, Edward Teller, Robert Oppenheimer, Hans Bethe, and Walter Brattain, with Claude Shannon, "the father of information theory," and with the statistician John Tukey. Hamming "became very interested in the difference between those who do and those who might have done," and he offered a number of answers to the question "why . . . so few scientists make significant contributions and so many are forgotten in the long run?" We have condensed Hamming's talk into the ten rules listed below:

## Rule 1: Drop Modesty

To quote Hamming: "Say to yourself: 'Yes, I would like to do first-class work.' Our society frowns on people who set out to do really good work. But you should say to yourself: 'Yes, I would like to do something significant.'"

## Rule 2: Prepare Your Mind

Many think that great science is the result of good luck, but luck is nothing but the marriage of opportunity and preparation. Hamming cites Pasteur's adage that "luck favours the prepared mind."

## Rule 3: Age Is Important

Einstein did things very early, and all the "quantum mechanic fellows," as well as most mathematicians and astrophysicists, were, as Hamming notes, "disgustingly young" when they did their best work. On the other hand, in the fields of music, politics, and literature, the protagonists often produce what we consider their best work late in life.

## Rule 4: Brains Are Not Enough, You Also Need Courage

Great scientists have more than just brainpower. To again cite Hamming: "Once you get your courage up and believe that you can do important things, then you can. If you think you can't, almost surely you are not going to. Great scientists will go forward under incredible circumstances; they think and continue to think."

## Rule 5: Make the Best of Your Working Conditions

To paraphrase Hamming, what most people think are the best working conditions clearly are not, because people are often most productive when working conditions are bad. One of the better times of the Cambridge Physical Laboratories was when they worked practically in shacks—they did some of the best physics ever. By turning the problem around a bit, great scientists often transform an apparent defect into an asset. "It is a poor workman who blames his tools—the good man gets on with the job, given what he's got, and gets the best answer he can."

## Rule 6: Work Hard and Effectively

Most great scientists have tremendous drive, and most of us would be surprised how much we would know if we worked as hard as some great scientists did for many years. As Hamming says: "Knowledge and productivity are like compound interest. Given two people with exactly the same ability, the one person who manages day in and day out to get in one more hour of thinking will be tremendously more productive over a lifetime." But, Hamming notes, hard work alone is not enough—it must be applied sensibly.

## Rule 7: Believe and Doubt Your Hypothesis at the Same Time

Great scientists tolerate ambiguity. They believe the theory enough to go

**Citation:** Erren TC, Cullen P, Erren M, Bourne PE (2007) Ten simple rules for doing your best research, according to Hamming. PLoS Comput Biol 3(10): e213. doi:10.1371/journal.pcbi.0030213



Thomas C. Erren is with the Institute and Policlinic for Occupational and Social Medicine, School of Medicine and Dentistry, University of Cologne, Köln, Lindenthal, Germany. Paul Cullen is with the Medizinisches Versorgungszentrum für Laboratoriumsmedizin Dr. Löer, Dr.Treder, Münster, Germany. Michael Erren is with the Institute of Clinical Chemistry and Laboratory Medicine, Westphalian Wilhelms-University of Münster, Münster, Germany. Philip E. Bourne is a Professor in the Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America.

*To whom correspondence should be addressed. E-mail: tim.erren@uni-koeln.de

ahead; they doubt it enough to notice the errors and faults so they can step forward and create the new replacement theory. As Hamming says: "When you find apparent flaws, you've got to be sensitive and keep track of those things, and keep an eye out for how they can be explained or how the theory can be changed to fit them. Those are often the great scientific contributions."

## Rule 8: Work on the Important Problems in Your Field

It is surprising but true that the average scientist spends almost all his time working on problems that he believes not to be important and not to be likely to lead to important results. By contrast, those seeking to do great work must ask: "What are the important problems of my field? What important problems am I working on?" Hamming again: "It's that simple. If you want to do great work, you clearly must work on important problems. . . . I finally adopted what I called 'Great Thoughts Time.' When I went to lunch Friday noon, I would only discuss great thoughts after that. By great thoughts I mean ones like: 'What will be the impact of computers on science and how can I change it?'"

## Rule 9: Be Committed to Your Problem

Scientists who are not fully committed to their problem seldom produce first-class work. To a large extent, creativity comes out of the subconscious. If you are deeply immersed in and committed to a topic, day after day, your subconscious has nothing to do but work on your problem. Hamming says it best: "So the way to manage yourself is that when you have a real important problem you don't let anything else get the center of your attention—you keep your thoughts on the problem. Keep your subconscious starved so it has to work on *your* problem, so you can sleep peacefully and get the answer in the morning, free."

## Rule 10: Leave Your Door Open

Keeping the door to your office closed makes you more productive in the short term. But ten years later, somehow you may not quite know what problems are worth working on, and all the hard work you do will be "sort of tangential" in importance. He (or she) who leaves the door open gets all kinds of interruptions, but he (or she) also occasionally gets clues as to what the world is and what might be important. Again, Hamming deserves to be quoted verbatim: "There is a pretty good correlation between those who work with the doors open and those who ultimately do important things, although people who work with doors closed often work harder. Somehow they seem to work on slightly the wrong thing—not much, but enough that they miss fame."

In our view, Rule 10 may be the key to getting the best research done because it will help you to obey Rules 1–9, and, most importantly, it will foster group creativity [9]. A discussion over lunch with your colleagues is often worth much more than a trip to the library. However, when choosing your lunchmates (and, by implication, your institution), be on your toes. As Hamming says: "When you talk to other people, you want to get rid of those sound absorbers who are nice people but merely say 'Oh yes,' and to find those who will stimulate you right back."


## Acknowledgments

**Funding.** The authors received no specific funding for this article.

**Competing interests.** The authors have declared that no competing interests exist.



**References**

1. Bourne PE (2005) Ten simple rules for getting published. PLoS Comp Biol 1: e57. doi:10.1371/journal.pcbi.0010057
2. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comp Biol 2: e12. doi:10.1371/journal.pcbi.0020012
3. Bourne PE, Korngreen A (2006) Ten simple rules for reviewers. PLoS Comp Biol 2: e110. doi:10.1371/journal.pcbi.0020110
4. Bourne PE, Friedberg I (2006) Ten simple rules for selecting a postdoctoral position. PLoS Comp Biol 2: e121. doi:10.1371/journal.pcbi.0020121
5. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comp Biol 3: e44. doi:10.1371/journal.pcbi.0030044
6. Bourne PE (2007) Ten simple rules for making good oral presentations. PLoS Comp Biol 3: e77. doi:10.1371/journal.pcbi.0030077
7. Erren TC, Bourne PE (2007) Ten simple rules for a good poster presentation. PLoS Comp Biol 3: e102. doi:10.1371/journal.pcbi.0030102
8. Hamming R (1986) You and your research. In: Kaiser JF Transcription of the Bell Communications Research Colloquium Seminar; 7 March 1986; Morristown, New Jersey, United States. Available: http://www.cs.virginia.edu/~robins/YouAndYourResearch.html. Accessed 24 September 2007.
9. Erren TC (2007) Hamming's "open doors" and group creativity as keys to scientific excellence: The example of Cambridge. Med Hypotheses 2007 Sep 3: 17804173.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for documenting scientific software

**Benjamin D. Lee** *

School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

* benjamindlee@me.com

## Introduction

Science, and especially biology, is increasingly relying on software tools to enable research. However, if you are a biologist, you likely received no training in software development best practices. Because of this lack of training, scientific software often has minimal or even nonexistent documentation, making the lives of researchers significantly harder than they need to be, with precious research time being spent figuring out how to use poorly documented software rather than performing the actual science. As the field matures, software documentation will become even more important as software stops being maintained and original authors are unable to be reached for support. Prior work has focused on various aspects of open software development [1–7], but documenting software has been underemphasized. I present these 10 simple rules in the hope that, by applying software engineering best practices to research tool documentation, you can create software that is maximally usable and impactful.

## Rule 1: Write comments as you code

Comments are the single most important aspect of software documentation. At the end of the day, people (yourself included) need to be able to read and understand your source code. Good variable and function names can help immensely with readability, although they are no complete replacement for comments. Although it may be perfectly obvious to you what your code does without comments, other readers will likely not be so fortunate. Indeed, you yourself may not even be able to understand your own code after you've moved on to another project. Think of comments as your lab notebook: they help you remember your train of thought long after the fact.

The best way to write comments is to do it as you code. That way you never have the problem of forgetting what your thought process was, and you never forget to go back and write the comments that you promised yourself you'd do (we're all guilty of this). Modern integrated development environments (IDEs) will often automatically generate documentation strings as you write code, which removes the burden of having to remember to write comments. One common argument against thorough code commenting is that it slows you down. In fact, good commenting can help you write code faster because you have a better understanding of your software. This understanding is especially useful when you run into bugs because you can compare what your code is doing to what your comments say it should be doing. Don't forget that, at the end of the day, your code has the final word on what your software will do, not your comments.

Proper code commenting is as much an art as it is a science. If you write too few comments, people won't be able to figure out what your code is doing. Write too many and readers will get lost in the sea of comments [4]. As a guiding principle, aim to write code that readers can

Check for updates

OPEN ACCESS

**Citation:** Lee BD (2018) Ten simple rules for documenting scientific software. PLoS Comput Biol 14(12): e1006561. https://doi.org/10.1371/journal.pcbi.1006561

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** December 20, 2018



**Funding:** The author received no specific funding for writing this article.

**Competing interests:** The author has declared that no competing interests exist.

understand purely by reading your comments [7]. If you remember one thing from this section, when in doubt, err on the side of more comments.

To get a feel for the right amount of commenting for code, let's examine some examples.

Bad (no comments):

```
for sequence in parsed_sequences:
        analyze(sequence)
```

Bad (too much commenting):

```
# iterate over the genes in the genome
for sequence in parsed_sequences:
        # call the analyze function, passing it each gene as
its argument
        analyze(sequence)
```

Good (just enough):

```
# analyze the genome
for sequence in parsed_sequences:
        analyze(sequence)
```

The key takeaway here is to keep your comments in the Goldilocks zone—not too many and not too few.

## Rule 2: Include examples (and lots of them)

When it comes to software documentation, showing takes precedence over telling. There are several important reasons to include examples in your documentation beyond simple instruction. Examples provide a starting point for experimentation. By starting from a piece of code that works, your users can attempt to change it for their own uses with minimal difficulty.

Unlike with comments, there isn't such a thing as too many examples if they all show off different aspects of your software. If you find that your main documentation is getting too laden with examples, feel free to move them to a special section or directory so long as you keep your examples organized and easily discoverable. Keras, a machine learning framework, has 35 full example scripts as of the time of this writing (github.com/keras-team/keras/tree/master/examples) with a README (see Rule 4 for more) explaining what each example demonstrates. Although you are by no means under any obligation to provide that many examples, do take the time to at least write examples showing off the main functionality of your software [2]. You can even make your examples do double duty as unit tests (or vice versa), thereby verifying functionality while providing instruction.

## Rule 3: Include a quickstart guide

Going from idea to experimentation to results as quickly as possible enables the progress of science. If people must spend a long time figuring out how to use your software, they're likely to give up. Conversely, if people can immediately start playing with your tool, they're vastly more likely to use it as a part of their research. It is therefore crucial to include a quickstart guide aimed at helping people begin using your software as quickly as possible.

This can take the form of an example (see Rule 2), a tutorial, a video, or anything else you can imagine. For example, let's look at the TPOT machine learning tool's quickstart guide [8]: it has an animated graphic image file (GIF) showing the software's functionality, diagrams explaining how it works, and a minimal code stub, perfect for copy-pasting into your own project. To tell whether your quickstart guide is working as intended, show it to someone who hasn't used your software and see if they can figure out how to start using it. Consider your quickstart guide to be a dating profile for your project: it should show off its strengths, give people a feel for it, and entice people into choosing it.

## Rule 4: Include a README file with basic information

Your README file acts like a homepage for your project. On code-sharing sites like GitHub, Bitbucket, and GitLab, your README file is shown on your project's main page. README files should be easily readable from the raw source, so human-readable markup languages such as Markdown or reStructuredText (or plain text) are preferable to less readable formats like hypertext markup language (HTML). In fact, code-sharing sites will usually render your markup language on your repository's page, giving you the best of both worlds. Take advantage of this—free hosting is hard to come by and the fact that your hosted README page is on your repository makes the arrangement even sweeter.

A good rule of thumb is to assume that the information contained within the README will be the only documentation your users read. For this reason, your README should include how to install and configure your software, where to find its full documentation, under what license it's released, how to test it to ensure functionality, and acknowledgments. Furthermore, you should include your quickstart guide (as introduced in Rule 3) in your README.

Often, the top of your README files will include badges that, when rendered, show the status of the software. One common source of badges is shields.io, which can dynamically generate badges for your project. Common badges include ones that show whether automated tests are passing (such those from travis-ci.org), what percentage of the code that the tests cover, whether the documentation is up to date, and more. Although not necessary, these badges instill confidence in the quality of your project and convey important information at a glance and are therefore highly recommended.

## Rule 5: Include a help command for command line interfaces

Many scientific software tools have command line interfaces (CLIs). Not having a graphical interface saves development time and makes the software more flexible. However, one challenge that CLI software has is that it can be hard to figure out how to use. The best way to document CLIs is to have a "help" command that will print out how to use the software. That way, users don't need to try to find your documentation to get basic tasks done. It should include usage (how to use the command), subcommands (if applicable), options and/or arguments, environment variables (if applicable), and maybe even some examples (Rule 2 strikes again!).

A help command can be tedious to make and difficult to maintain, but luckily there are numerous software packages that can do it for you. In Python, software such as click (click.pocoo.org) can not only make your help command but can also even help you make your interface, saving you time and effort.

An example of a good CLI is the one included in Magic-BLAST. It has a short help command, "-h," which provides basic information on what the tool is and how to use it. It also includes instructions on how to access the full help documentation, which include a list of every option as well a description of the option's arguments and what it does. An arrangement like this is particularly good because it requires minimal effort to find just the most useful information via the short help page, thereby reducing information overload and reducing the cognitive load of using the software by providing a reminder of how to access the full CLI reference.

## Rule 6: Version control your documentation

A previous Ten Simple Rules article has described the virtues of using Git for your code [1]. Because your documentation is such an integral part of your code, it must be version controlled as well. To start, you should keep your documentation inside your Git repository along with the rest of your files. This makes it possible to view your documentation at any point in the project's history. Services such as Read the Docs (readthedocs.org) and Zenodo (zenodo.org) make doing this even easier because they will archive a complete rendered version of your documentation every time you make a new release of your software.

To illustrate why this is such an important rule, consider what would happen if you change a default setting in a new release of your software. When users of previous versions go to look at your documentation, they will see the documentation that is incompatible with the version that they have installed. Worse still, because you changed a default, the software could fail inexplicably. This can be incredibly aggravating to users (and even dangerous if the software is for mission-critical applications), so it is extra important to use version control for your documentation. A changelog in your documentation can make this task much easier. If you are using informative commit messages, creating a changelog is a straightforward task at worst and a trivial task at best.

As an example of a bioinformatics library that is doing a particularly good job at version controlling their documentation, look at khmer, which has a thorough changelog containing new features, fixed bugs (separated by whether they are relevant to users or developers), known issues, and a list of the contributors to the release [9]. In addition, previous versions of the documentation website are easily accessible and labeled clearly. By providing this information, the authors have ensured that users of any version of the software can get the right version of the documentation, see what's going on in the project, and make sure they're aware of any issues with their version.

If you take one thing away from this rule, make it very clear which version of your software your documentation is for and preserve previous versions of your documentation—your users will thank you.

## Rule 7: Fully document your application programming interface

Your application programming interface (API) is how people who are using your software interact with your code. It is imperative that it be fully documented in the source code. In all honesty, probably nobody will read your entire API documentation, and that's perfectly fine. The goal of API documentation is to prevent users from having to dig into your (well commented, right?) source code to use your API. At the very least, each function should have its inputs and input types noted, its output and output type noted, and any errors it can raise documented. Objects should have their methods and attributes described. It's best to use a consistent style for your API documentation. The Google style guide (google.github.io/styleguide) has API documentation suggestions for numerous languages such as Python, Java, R, C++, and Shell. You spent a lot of time developing your API; don't let that time go to waste by not telling your users how to use it.

## Rule 8: Use automated documentation tools

The best type of documentation is documentation that writes itself. Although no software package can do all your documentation for you (yet), there are tools that can do much of the heavy lifting, such as making a website, keeping it in sync with your code, and rendering it to a portable document file (PDF). Software such as Sphinx (sphinx-doc.org), perldoc, Javadoc, and Roxygen (https://github.com/klutometis/roxygen) for R can generate documentation and even read your comments and use those to generate detailed API documentation. Although Sphinx was developed to host Python's documentation, it is language agnostic, meaning that it can work for whatever language your project is in. Similarly, Doxygen (doxygen.nl) and MkDocs (mkdocs.org) are language-agnostic documentation tools. Read the Docs, introduced in Rule 6, is a language-agnostic documentation hosting platform that can rebuild your documentation every time that you push to your repository, ensuring that your documentation is always up to date.

There are many other ways automation can make your documentation smarter: in Python, software like doctest (sphinx-doc.org/en/stable/ext/doctest.html) can automatically pull examples from your documentation and ensure that your code does what you say it should be doing in the documentation. To help you follow Rule 7, there are tools such as Napoleon (github.com/sphinx-contrib/napoleon) that can generate your API documentation for you. It's even possible to automatically generate interactive representational state transfer (REST) API documentation using free tools such as Swagger (swagger.io). At this point, there is almost no reason not to be using automated documentation tools.

## Rule 9: Write error messages that provide solutions or point to your documentation

Error messages are part of life when developing software. As a developer, you should be doing your best to make your error messages as informative as possible. Good error messages should have three parts: they should state what the error is, what the state of the software was when it generated the error, and either how to fix it or where to find information relevant to fixing it. In the spirit of Rule 2, let's look at an example.

Bad:

```
Error: Translation failed.
```

Good:

```
Error: Translation failed because of an invalid codon ("IQT")
in position 1 in sequence 41. Ensure that this is a valid DNA
sequence and not a protein sequence.
```

By showing what exactly went wrong and proposing a fix for it, your users will spend less time debugging and more time doing science. Since you know your software better than anyone else, providing guidance in error messages can be invaluable. If for no other reason, do it to save yourself the hassle of being tech support for your users (most of whom have barely read your documentation, if at all) when they run into easily fixable usage mistakes.

Furthermore, it is important to say what the state of the software was when the error was generated, especially if it takes a long time to run or you don't save logs by default. If your software fails, seemingly at random, after 12 hours of execution, your users will be thankful to know what was going on when the error was thrown rather than having to wait another 12 hours to reproduce the error with logging enabled.

## Rule 10: Tell people how to cite your software

Of all the rules in this guide, odds are that this is the one you need the least. However, it must be said that, if you publish scientific software, you need to include the information required to properly provide attribution to your work. I recommend providing the digital object identifier (DOI), a BibTeX entry, and a written reference for your publication in your README, as well as using a "CITATION" file in citation file format (CFF) format, which is a human- and machine-readable file format designed for specifying citation information for scientific software [10].

Including citation information in your documentation is especially important for software that has not been published in a traditional academic journal, which would assign it a DOI. Just because your software is unpublished doesn't mean that you can't get a DOI for it—you deserve credit for your work. If you're using Zenodo to archive your releases (see Rule 6), it will mint a new DOI for each release as well as a DOI for the entire project. Another great, free way to get a DOI for your project is to submit it to the Journal of Open Source Software (joss.theoj.org), a peer-reviewed open-access academic journal designed for software developers. Both even provide a badge for your README (see Rule 4) so that the entire world can tell how to cite your software at a glance.

## Conclusion

I hope that this guide will help you improve the quality of your software documentation. Documenting software is not always as exciting as doing original software development, but it is nonetheless as important. Software documentation is in many ways like writing a paper in that it is a required step in the dissemination of your ideas. It is a critical step for ensuring reproducibility, not to mention the fact that many bioinformatics journals now require that software submitted be well documented. Automated documentation tools such as Sphinx can diminish some of the effort required for good documentation, perhaps even making the work enjoyable. Finally, because documentation can make or break a project's adoption in the real world, by following these 10 simple rules you can give your project its best chance of wide adoption and possibly even end up as an example of good documentation in a Ten Simple Rules article!

## Acknowledgments

I would like to thank Ben Busby for his feedback on this manuscript.

## References

1. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost F da V, et al. Ten simple rules for taking advantage of git and GitHub. PLoS Comput Biol. 2016; https://doi.org/10.1371/journal.pcbi.1004947 PMID: 27415786
2. List M, Ebert P, Albrecht F. Ten simple rules for developing usable software in computational biology. PLoS Comput Biol. 2017; https://doi.org/10.1371/journal.pcbi.1005265 PMID: 28056032
3. Prlić A, Procter JB. Ten simple rules for the open development of scientific software. PLoS Comput Biol. 2012; https://doi.org/10.1371/journal.pcbi.1002802 PMID: 23236269
4. Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. Best practices for scientific computing. PLoS Biol. 2014; https://doi.org/10.1371/journal.pbio.1001745 PMID: 24415924
5. Blischak JD, Davenport ER, Wilson G. A quick introduction to version control with git and GitHub. PLoS Comput Biol. 2016; https://doi.org/10.1371/journal.pcbi.1004668 PMID: 26785377
6. Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, Dalchau N, et al. Ten simple rules for effective computational research. PLoS Comput Biol. 2014; https://doi.org/10.1371/journal.pcbi.1003506 PMID: 24675742
7. Noble WS. A quick guide to organizing computational biology projects. PLoS Comput Biol. 2009; https://doi.org/10.1371/journal.pcbi.1000424 PMID: 19649301
8. Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. Applications of Evolutionary Computation; 2016. https://doi.org/10.1007/978-3-319-31204-0_9
9. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, et al. The khmer software package: enabling efficient nucleotide sequence analysis. F1000Research; 2015. https://doi.org/10.12688/f1000research.6924.1 PMID: 26535114
10. Druskat S. Citation File Format (CFF); 2013. Database: zenodo [Internet]. Available from: https://zenodo.org/record/1242911. [cited 2018 Jul 17]. https://doi.org/10.5281/zenodo.1242911

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for biologists learning to program

**Maureen A. Carey**[1], **Jason A. Papin**[2]*

**1** Department of Microbiology, Immunology, and Cancer Biology, University of Virginia School of Medicine, Charlottesville, Virginia, United States of America, **2** Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, United States of America

* papin@virginia.edu





**Citation:** Carey MA, Papin JA (2018) Ten simple rules for biologists learning to program. PLoS Comput Biol 14(1): e1005871. https://doi.org/10.1371/journal.pcbi.1005871

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** January 4, 2018



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

Jason A. Papin is co-Editor-in-Chief of *PLOS Computational Biology*.

## Introduction

As big data and multi-omics analyses are becoming mainstream, computational proficiency and literacy are essential skills in a biologist's tool kit. All "omics" studies require computational biology: the implementation of analyses requires programming skills, while experimental design and interpretation require a solid understanding of the analytical approach. While academic cores, commercial services, and collaborations can aid in the implementation of analyses, the computational literacy required to design and interpret omics studies cannot be replaced or supplemented. However, many biologists are only trained in experimental techniques. We write these 10 simple rules for traditionally trained biologists, particularly graduate students interested in acquiring a computational skill set.

## Rule 1: Begin with the end in mind

When picking your first language, focus on your goal. Do you want to become a programmer? Do you want to design bioinformatic tools? Do you want to implement tools? Do you want to just get these data analyzed already? Pick an approach and language that fits your long- and short-term goals.

Languages vary in intent and usage. Each language and package was created to solve a particular problem, so there is no universal "best" language (Fig 1). Pick the right tool for the job by choosing a language that is well suited for the biological questions you want to ask. If many people in your field use a language, it likely works well for the types of problems you will encounter. If people in your field use a variety of languages, you have options. To evaluate ease of use, consider how much community support a language has and how many resources that community has created, such as prevalence of user development, package support (documentation and tutorials), and the language's "presence" on help pages. Practically, languages vary in cost for academic and commercial use. Free languages are more amenable to open source work (i.e., sharing your analyses or packages). See Table 1 for a brief discussion of several programming languages, their key features, and where to learn more.

## Rule 2: Baby steps are steps

Once you've begun, focus on one task at a time and apply your critical thinking and problem solving skills. This requires breaking a problem down into steps. Analyzing omics data may sound challenging, but the individual steps do not: e.g., read your data, decide how to interpret missing values, scale as needed, identify comparison conditions, divide to calculate fold change, calculate significance, correct for multiple testing. Break a large problem into modular

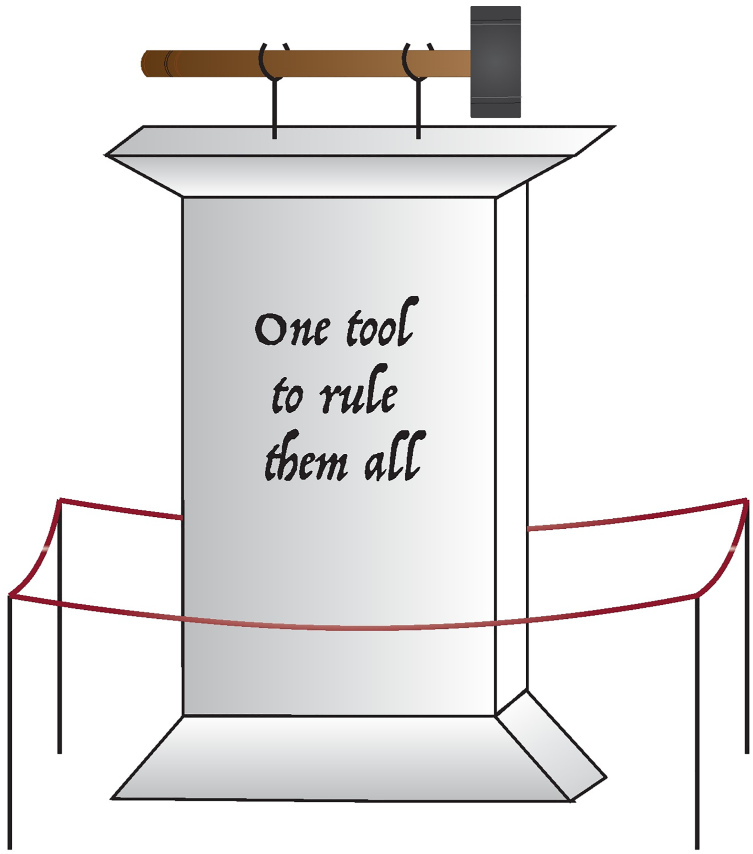



**Fig 1. The "one tool to rule them all" (or: how programming languages do not work).**

https://doi.org/10.1371/journal.pcbi.1005871.g001

tasks and implement one task at a time. Iteratively edit for efficiency, flow, and succinctness. Mistakes will happen. That's ok; what matters is that you find, correct, and learn from them.

## Rule 3: Immersion is the best learning tool

Don't stitch together an analysis by switching between or among languages and/or point and click environments (Excel [Microsoft; https://www.microsoft.com/en-us/], etc.). While learning, if a job can be done in one language or environment, do it all there. For example, importing a spreadsheet of data (like you would view in Excel) is not necessarily straightforward; Excel automatically determines how to read text, but the method may differ from conventions in other programming languages. If the import process "misreads" your data (e.g., blank cells are not read as blank or "NA," numbers are in quotes indicating that they are read as text, or column names are not maintained), it can be tempting to return to Excel to fix these with search-and-replace strategies. However, these problems can be fixed by correctly reading the data and by understanding the language's data structures. Just like a spoken language [1, 2], immersion is the best learning tool [3, 4]. In addition to slowing the learning curve, transferring across programs induces error. See References [5–7] for additional Excel or word processing–induced errors.

Eventually, you may identify tasks that are not well suited to the language you use. At that point, it may be helpful to pick up another language in order to use the right tool for the job

**Table 1. A noninclusive discussion of programming languages.** A **shell** is a command line (i.e., programming) interface to an operating system, like **Unix** operating systems. **Low-level** programming languages deal with a computer's hardware. The process of moving from the literal processor instructions toward human-readable applications is called "abstraction." Low-level languages require little abstraction. **Interpreted** languages are quicker to test (e.g., to run a few lines of code); this facilitates learning through trial and error. Interpreted languages tend to be more human readable. **Compiled** languages are powerful because they are often more efficient and can be used for low-level tasks. However, the distinction between interpreted and compiled languages is not always rigid. All languages presented below are free unless noted otherwise. The Wikipedia page on programming languages provides a great overview and comparison of languages.

| Language | Key features | Documentation | Sample tutorials | Community groups |
|---|---|---|---|---|
| **Bash** | • Most common Unix shell<br>• Practical for execution of scripts written in all other languages<br>• Versatile<br>• Easy to delete files or make other drastic changes<br>• Weaknesses include executing math and limited data structures<br>• Default for macOS and most Linux distributions | • gnu.org/software/bash/manual/<br>• On macOS's terminal, type "man <command>" to get the manual for any command (and "q" to exit manual page) | • The Linux Documentation Project's Beginner's guide: tldp.org/LDP/Bash-Beginners-Guide/html/<br>• Ubuntu's documentation: help.ubuntu.com/community/Beginners/BashScripting<br>• Azet's GitHub page: github.com/azet/community_bash_style_guide | • Google Plus: plus.google.com/communities/110832059019676429606<br>• GitHub community resources page: github.com/awesome-lists/awesome-bash |
| **Python** | • General purpose language<br>• Considered easy to learn due to readability<br>• Flexible syntax considered both a strength and weakness<br>• Interpreted language | • docs.python.org | • Google's Python class: developers.google.com/edu/python/<br>• The Hitchhiker's Guide to Python: docs.python-guide.org/ | • Python Users Group: wiki.python.org/moin/LocalUserGroups<br>• Python Special Interest Groups: python.org/community/sigs/ |
| **R** | • Community involvement<br>• Application-focused development<br>• Easy to learn by coupling basic programming and applications<br>• Well-developed visualization<br>• Variable package quality<br>• "Tidy data" community<br>• Interpreted language | • rdocumentation.org<br>• r-project.org<br>• cran.r-project.org | • R for cats: rforcats.net<br>• Books by Hadley Wickham: hadley.nz<br>• R Tutorial's introduction: r-tutor.com/r-introduction<br>• Cyclismo's R Tutorial: cyclismo.org/tutorial/R/ | • R-Ladies: rladies.org<br>• R Users Group: many |
| **SAS** | • Statistical computing<br>• High-quality development of statistical functions by commercial and academic developers<br>• Domain-specific usage<br>• Free for students only<br>• Typically a compiled language | • support.sas.com | • Boston University's SAS Training for Statistics: bu.edu/stat/bu-student-chapter-of-the-asa/sas-training/ | • SAS User Groups: sas.com/en_us/connect/user-groups.html |
| **MATLAB** | • Well-developed applications in engineering<br>• Maintained professionally<br>• Interpreted language<br>• Discounted academic license | • mathworks.com/help/matlab | • Cyclismo's MATLAB Tutorial: cyclismo.org/tutorial/matlab/<br>• For purchase courses offered at: matlabacademy.mathworks.com | • MATLAB Central: mathworks.com/matlabcentral/ |
| **Perl** | • General purpose language<br>• Handles text well<br>• Waning community involvement<br>• Syntax modelled after human language<br>• Interpreted language | • perl.org<br>• cpan.org | • Beginning Perl: perl.org/books/beginning-perl/<br>• Perl maven's tutorial: perlmaven.com<br>• Perl::Learn: learn.perl.org | • Perl Mongers: pm.org<br>• Perl Monks: perlmonks.org |

(*Continued*)

**Table 1.** (*Continued*)

| Language | Key features | Documentation | Sample tutorials | Community groups |
|---|---|---|---|---|
| **Fortran** | • Numeric computation<br>• Fast<br>• Often used for high-performance computing<br>• Limited development<br>• Compiled language | • fortranwiki.org | • many at Fortran wiki: fortranwiki.org/fortran/show/Tutorials | • Fortran Friends: fortran.orpheusweb.co.uk |
| **C/C++** | • Low-level language<br>• Powerful, used for source code of many other languages<br>• Challenging to learn as it requires explicit syntax<br>• Explicit syntax enforces good programming habits<br>• Compiled language | • devdocs.io/c<br>• cppreference.com | • C programming's tutorial: cprogramming.com/tutorial/<br>• Learn-C's web-based tutorial: learn-c.org | • Standard C++ Foundation: isocpp.org<br>• C/C++ Users Group (CUG): hal9k.com/cug |

https://doi.org/10.1371/journal.pcbi.1005871.t001

(see Rule 1). In fact, understanding one language will make it easier to learn a second. Until then, however, focus on immersion to learn.

## Rule 4: Phone a friend

There are numerous online resources: tutorials, documentation, and sites intended for community Q and A (StackOverflow, StackExchange, Biostars, etc.), but nothing replaces a friend or colleague's help. Find a community of programmers, ranging from beginning to experienced users, to ask for help. You may want to look for both technical support (i.e., a group centered around a language) and support regarding a particular scientific application (e.g., a group centered around omics analyses). Many universities have scientific computing groups, housed in the library or information technology (IT) department; these groups can be your starting point. If your lab or university does not have a community of programmers, seek them out virtually or locally. Coursera courses, for example, have comment boards for students to answer each other's questions and learn from their peers. Organizations like Software and Data Carpentry or language user groups have mailing lists to connect members. Many cities have events organized by language-specific user groups or interest groups focused on big data, machine learning, or data visualization. These can be found through meetup.com, Google groups, or through a user group's website; some are included in Table 1.

Once you find a community, ask for help. At the beginning stages, in-person help to deconstruct or interpret an online answer is invaluable. Additionally, ask a friend for code. You wouldn't write a paper without first reading a lot of papers or begin a new project without shadowing a few experimenters. First, read their code. Implement and interpret, trying to understand each line. Return to discuss your questions. Once you begin writing, ask for edits.

## Rule 5: Learn how to ask questions

There's an answer to almost anything online, but you have to know what to ask to get help. In order to know what to ask, you have to understand the problem. Start by interpreting an error message. Watch for generic errors and learn from them. Identify which component of your error message indicates what the issue is and which component indicates where the issue is (Figs 2–5). Understanding the problem is essential; this process is called "debugging." Without truly understanding the problem, any "solution" will ultimately propagate and escalate the mistake, making harder-to-interpret errors down the road. Once you understand the problem,

| Code (input and output) | Debugging approach |
|---|---|
| > a = 1<br>> b = 'dogs'<br>> c = 2 | |
| > # do math ( goal: a + c )<br>> d = a + c<br>> d<br>[1] 3 | |
| > # combine two variables ( goal: '3 dogs' )<br>> d + b<br>Error in d + b : non-numeric argument to binary operator | Goal: Here we aim to combine two strings of text into a longer string.<br>Problem: 'dogs' is not a number, and "+" requires numeric inputs. |
| > # first make '3' (variable d) a word, or a 'string'<br>> d = toString(d)<br>> d<br>[1] "3" | Debugging Step: Try googling 'R combine two words.' |
| > paste(d,b)<br>[1] "3 dogs" | Lesson learned: Numbers can be stored as numeric or string variables, and functions require specific input types. |

**Fig 2. Anatomy of an error message, Part 1 (or: How to write more than one line of code).** Here we show an example of the debugging process in R using the RStudio environment, with the goal of concatenating two words.

https://doi.org/10.1371/journal.pcbi.1005871.g002

look for answers. Looking for answers requires effective googling. Learn the vocabulary (and meta-vocabulary) of the language and its users. Once you understand the problem and have identified that there is no obvious (and publicly available) solution, ask for answers in programming communities (see Rule 4 and Table 1). When asking, paraphrase the fundamental problem. Include error messages and enough information to reproduce the problem (include packages, versions, data or sample data, code, etc.). Present a brief summary of what was done, what was intended, how you interpret the problem, what troubleshooting steps were already taken, and whether you have searched other posts for the answer.

See the following website for suggestions: http://codereview.stackexchange.com/help/how-to-ask and [8]. End with a "thank you" and wait for the help to arrive.

## Rule 6: Don't reinvent the wheel

Rule 6 can also be found in "Ten Simple Rules for the Open Development of Scientific Software" [9], "Ten Simple Rules for Developing Public Biological Databases" [10], "Ten Simple Rules for Cultivating Open Science and Collaborative R&D" [11], and "Ten Simple Rules To Combine Teaching and Research" [12]. Use all resources available to you, including online tutorials, examples in the language's documentation, published code, cool snippets of code your labmate shared, and, yes, your own work. Read widely to identify these resources. Copy-and-paste is your friend. Provide credit if appropriate (i.e., comment "adapted from so-n-so's X script") or necessary (e.g., read through details on software licenses). Document your scripts by commenting in notes to yourself so that you can use old code as a template for future work.

| Code (input and output) | Debugging approach |
|---|---|

In[1]:

```
# load the 'pandas' package, which allows us to use the data structure called 'DataFrame'
import pandas
# load dataset
df = pandas.read_csv('http://bioconnector.org/workshops/data/gapminder.csv')
# print top 5 rows of our dataset
print df.head()
```

Out[1]:

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|---|---|
| 0 | Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.445314 |
| 1 | Afghanistan | Asia | 1957 | 30.332 | 9240934 | 820.853030 |
| 2 | Afghanistan | Asia | 1962 | 31.997 | 10267083 | 853.100710 |
| 3 | Afghanistan | Asia | 1967 | 34.020 | 11537966 | 836.197138 |
| 4 | Afghanistan | Asia | 1972 | 36.088 | 13079460 | 739.981106 |

In[2]:

```
# let's extract only the data regarding countries in Africa
df['continent'] == 'Africa'
```

Out[2]:

```
0        False
1        False
2        False
   ...
1702     True
1703     True
```

Goal: Here we aim to extract the data associated with countries in Africa. We want to subset the dataframe to extract these datapoints.

Hint 1: If your command can be executed, there will not be an error message. Check output to see if your results are what you aim to get.

In[3]:

```
# let's use the boolean statement to subset the dataframe
df.loc[(df['continent'] == 'Africa')]
df.head()
```

Out[3]:

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|---|---|
| 24 | Algeria | Africa | 1952 | 43.077 | 9279525 | 2449.008185 |
| 25 | Algeria | Africa | 1957 | 45.685 | 10270856 | 3013.986023 |
| 26 | Algeria | Africa | 1962 | 48.303 | 11000948 | 2550.816880 |
| 27 | Algeria | Africa | 1967 | 51.407 | 12760499 | 3256.991771 |
| 28 | Algeria | Africa | 1972 | 54.518 | 14760787 | 4182.663766 |

Problem 1: We identified which datapoints we want, but did not extract the data.

Debugging step 1: Google 'python pandas selecting data,' and read documentation.

Lesson learned: View output and compare to your expected results to identify 'silent' errors.

**Fig 3. Anatomy of an error message, Part 2 (or: Just because it works, doesn't mean it's right).** Here we provide more examples of the debugging process. Examples shown in Figs 3–5 are conducted in Python using a Jupyter notebook. Environments like RStudio (in Fig 2) and Jupyter notebooks are two examples of integrated development environments; these environments offer additional support, including built-in debugging tools. First, we show an error that does not induce an error message, but the user must debug nonetheless.

https://doi.org/10.1371/journal.pcbi.1005871.g003

These comments will help you remember what each line of code intends to do, accelerating your ability to find mistakes.

## Rule 7: Develop good habits early on

Computational research is research, so use your best practices. This includes maintaining a computational lab notebook and documenting your code. A computational lab notebook is by definition a lab notebook: your lab notebook includes protocols, so your computational lab notebook should include protocols, too. Computational protocols are scripts, and these should include the code itself and how to access everything needed to implement the code. Include

| Code (input and output) | Debugging approach |
|---|---|
| | |

In[1]:

```
# load package and dataset
import pandas
df = pandas.read_csv('http://bioconnector.org/workshops/data/
gapminder.csv')
print df.head()
```

Out[1]:

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|---|---|
| 0 | Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.445314 |
| 1 | Afghanistan | Asia | 1957 | 30.332 | 9240934 | 820.853030 |
| 2 | Afghanistan | Asia | 1962 | 31.997 | 10267083 | 853.100710 |
| 3 | Afghanistan | Asia | 1967 | 34.020 | 11537966 | 836.197138 |
| 4 | Afghanistan | Asia | 1972 | 36.088 | 13079460 | 739.981106 |

In[2]:

```
# now let's see how many countries are represented in the data
df['Country'].count()
```

Goal: Here we aim to count how many countries are represented in the dataset.

Out[2]:

```
KeyError                                  Traceback (most recent call last)

<ipython-input-4-2e5d10c97161> in <module>()
          1 # now let's see how many countries are represented in the data
---->     2 df['Country'].count()

/Users/Maureen/.virtualenvs/python2.7/site-packages/pandas/core/frame.pyc in
__getitem__(self,key)
      1962                return self._getitem_multilevel(key)
      1963          else:
      1964                return self._getitem_column(key)
      1965
----> 1966      def _getitem_column(self,key):

/Users/Maureen/.virtualenvs/python2.7/site-packages/pandas/core/frame.pyc in
_getitem_column(self,key)
      1969          # get column
      1970          if self.columns.is_unique:
----> 1971                return self._get_item_cache(key)
      1972
      1973          # duplicate columns & possible reduce dimensionality

                                   ...
KeyError: 'Country'
```

Hint 2a: A traceback is a sequence of calls that lead to an error.

Problem 2: Error message reports an error with the string `Country` in line 2.

Hint 2b: None of the errors shown involve 'count.'

Hint 2c: All functions involve 'self,' which is the original dataframe, and a 'key,' which is the search term. Maybe something is wrong with the search term, 'Country.'

Debugging step 2: Try googling: 'python pandas call dataframe by column' to see examples of implementation.
Also, double check that 'Country' is a column. (See output 1 above.)

In[3]:

```
df['country'].count()
```

Out[3]: 1704

Lesson learned: Many functions are spelling and case sensitive.

**Fig 4. Anatomy of an error message, Part 3 (or: Trace your way back to the problem).** Here we show an explicit error message.

https://doi.org/10.1371/journal.pcbi.1005871.g004

input (raw data) and output (results), too. Figures and interpretation can be included if that's how you organize your lab notebook. Develop computational "place habits" (file-saving strategies). It is easier to organize one drawer than it is to organize a whole lab, so start as soon as you begin to learn to program. If you can find that experiment you did on June 12, 2011—its protocol and results—in under five minutes, you should be able to find that figure you generated for lab meeting three weeks ago, complete with code and data, in under five minutes as well. This requires good version control or documentation of your work. Like with protocols,

| Code (input and output) | Debugging approach |
|---|---|

In[1]:
```
# load package and dataset
import pandas
df = pandas.read_csv('http://bioconnector.org/workshops/data/gapminder.csv')
```

In[2]:
```
# let's see if the mean life expectancy (lifeExp) changes by year
new_df = df.groupby('year',as_index = False)['lifeExp'].mean()
new_df.head()
```

Goal: Here we aim to find the mean life expectancy each year. We found the code by googling 'python pandas find mean of groups,' but we do not understand the line of code that provides this result. Thus, we will debug the line to understand it.

Out[2]:

| | year | lifeExp |
|---|---|---|
| 0 | 1952 | 49.057620 |
| 1 | 1957 | 51.507401 |
| 2 | 1962 | 53.609249 |
| 3 | 1967 | 55.678290 |
| 4 | 1972 | 57.647386 |

In[3]:
```
# We found a solution, but what does it mean? Let's break it down.
# df.groupby: the function groupby is applied to the dataframe df
# groupby requires a grouping variable ('year')
# Next we take the mean of column 'lifeExp' of this grouped df
# but then what does the 'as_index = False' do?
# let's try changing 'False' to 'True' to see how the results change
new_df = df.groupby('year',as_index = True)['lifeExp'].mean()
new_df.head()
```

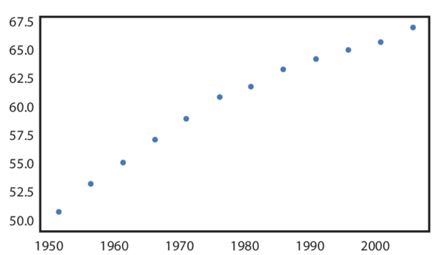
Debugging step 3: Break down each component of code into knowns and unknowns. Use documentation and experimentation to understand your unknowns.

Out[3]:
```
year
1952   49.057620
1957   51.507401
1962   53.609249
1967   55.678290
1972   57.647386
Name: lifeExp, dtype: float64
```

Hint 3b: The dataframe's index is an axis label. Try googling 'python pandas index' to learn more. The new output is fine, unless you want both variables (year and mean life expectancy) as columns. We want each variable in a column to make plotting easier.

In[4]:
```
# if we want both variables as columns, we use 'as_index = False'
plot_df = df.groupby('year',as_index = False)['lifeExp'].mean()
# to make a plot of these data, first load a plotting package. We
can specify the shorthand 'plt' to make accessing functions easier
import matplotlib.pyplot as plt
# make plot by setting plot type (scatter) and x and y variables
plt.scatter(x = plot_df['year'], y = plot_df['lifeExp'])
plt.show()
```

Out[4]:




Lesson learned: Debug to understand your solution, not just the problem.

**Fig 5. Anatomy of an error message, Part 4 (or: Debugging a solution).** Lastly, we show how to debug a solution to understand a line of code found on the internet.

https://doi.org/10.1371/journal.pcbi.1005871.g005

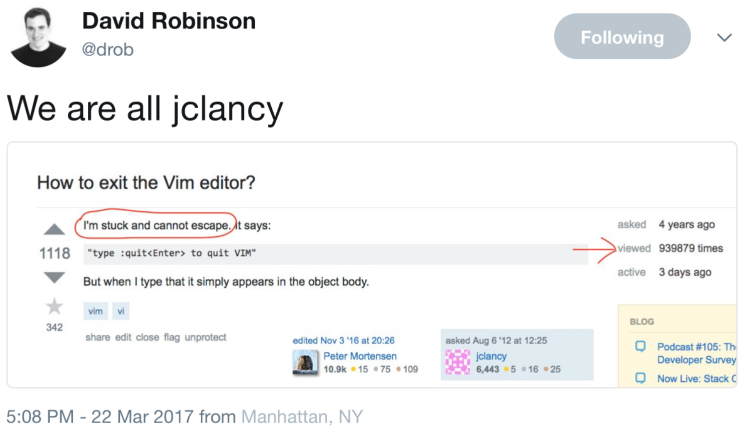



**Fig 6. "How to exit the vim editor?" (or: We all get stuck at some point).** Now viewed >1.33 million times; see: http://stackoverflow.com/questions/11828270/how-to-exit-the-vim-editor.

https://doi.org/10.1371/journal.pcbi.1005871.g006

each time you run a script, you should note any modifications that are made. Document all changes in experimental and computational protocols. These habits will make you more efficient by enhancing your work's reproducibility. For specific advice, see "Ten Simple Rules for a Computational Biologist's Laboratory Notebook" [13], "Ten Simple Rules for Reproducible Computational Research" [14], and "Ten Simple Rules for Taking Advantage of Git and GitHub" [15].

## Rule 8: Practice makes perfect

Use toy datasets to practice a problem or analysis. Biological data get big, fast. It's hard to find the computational needle-in-a-haystack, so set yourself up to succeed by practicing in controlled environments with simpler examples. Generate small toy datasets that use the same structure as your data. Make the toy data simple enough to predict how the numbers, text, etc., should react in your analysis. Test to ensure they do react as expected. This will help you understand what is being done in each step and troubleshoot errors, preparing you to scale up to large, unpredictable datasets. Use these datasets to test your approach, your implementation, and your interpretation. Toy datasets are your negative control, allowing you to differentiate between negative results and simulation failure.

## Rule 9: Teach yourself

How would you teach you if you were another person? You would teach with a little more patience and a bit more empathy than you are practicing now. You are not alone in your occasional frustration (Fig 6). Learning takes time, so plan accordingly. Introductory courses are helpful to learn the basics because the basics are easy to neglect in self-study. Articulate clear expectations for yourself and benchmarks for success. Apply some of the structure (deadlines, assignments, etc.) you would provide a student to help motivate and evaluate your progress. If something isn't working, adjust; not everyone learns best by any one approach. Explore tutorials, online classes, workshops, books like *Practical Computing for Biologists* [16], local programming meetups, etc., to find your preferred approach.

## Rule 10: Just do it

Just start coding. You can't edit a blank page.

Learning to program can be intimidating. The power and freedom provided in conducting your own computational analyses bring many decisions points, and each decision brings more room for mistakes. Furthermore, evaluating your work is less black-and-white than for some experiments. However, coding has the benefit that failure is risk free. No resources are wasted —not money, time (a student's job is to learn!), or a scientific reputation. In silico, the playing field is leveled by hard work and conscientiousness. So, while programming can be intimidating, the most intimidating step is starting.

## Conclusion

Markowetz recently wrote, "Computational biologists are just biologists using a different tool" [17]. If you are a traditionally trained biologist, we intend these 10 simple rules as instruction (and pep talk) to learn a new, powerful, and exciting tool. The learning curve can be steep; however, the effort will pay dividends. Computational experience will make you more marketable as a scientist (see "Top N Reasons To Do A Ph.D. or Post-Doc in Bioinformatics/Computational Biology" [18]). Computational research has fewer overhead costs and reduces the barrier to entry in transitioning fields [19], opening career doors to interested researchers. Perhaps most importantly, programming skills will make you better able to implement and interpret your own analyses and understand and respect analytical biases, making you a better experimentalist as well. Therefore, the time you spend at your computer is valuable. Acquiring programming expertise will make you a better biologist.

## Acknowledgments

Thank you to Ed Hall, Pat Schloss, Matthew Jenior, Angela Zeigler, Jhansi Leslie, and Gregory Medlock for their feedback.

## References

1. Genesee F. Integrating language and content: Lessons from immersion. Center for Research on Education, Diversity & Excellence. 1994.
2. Genesee FH, editor Second language learning in school settings: Lessons from immersion1991: Lawrence Erlbaum Associates.
3. Campbell W, Bolker E, editors. Teaching programming by immersion, reading and writing2002: IEEE.
4. Guzdial M. Programming environments for novices. Computer science education research. 2004; 2004:127–54.
5. Zeeberg BR, Riss J, Kane DW, Bussey KJ, Uchio E, Linehan WM, et al. Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics. BMC Bioinformatics. 2004; 5(1):80.
6. Ziemann M, Eren Y, El-Osta A. Gene name errors are widespread in the scientific literature. Genome Biol. 2016; 17(1):177. https://doi.org/10.1186/s13059-016-1044-7 PMID: 27552985
7. Linke D. Commentary: Never trust your word processor. Biochemistry and Molecular Biology Education. 2009; 37(6):377–. https://doi.org/10.1002/bmb.20340 PMID: 21567776
8. Collado-Torres L. Recent Posts [Internet]2017. [cited 2017]. Available from: http://lcolladotor.github.io/. Posts. Accessed on 5 April 2017.
9. Prlić A, Procter JB. Ten simple rules for the open development of scientific software. PLoS Comput Biol. 2012; 8(12):e1002802. https://doi.org/10.1371/journal.pcbi.1002802 PMID: 23236269
10. Helmy M, Crits-Christoph A, Bader GD. Ten Simple Rules for Developing Public Biological Databases. PLoS Comput Biol. 2016; 12(11):e1005128. https://doi.org/10.1371/journal.pcbi.1005128 PMID: 27832061
11. Masum H, Rao A, Good BM, Todd MH, Edwards AM, Chan L, et al. Ten simple rules for cultivating open science and collaborative R&D. PLoS Comput Biol. 2013; 9(9):e1003244. https://doi.org/10.1371/journal.pcbi.1003244 PMID: 24086123
12. Vicens Q, Bourne PE. Ten simple rules to combine teaching and research. PLoS Comput Biol. 2009; 5 (4):e1000358. https://doi.org/10.1371/journal.pcbi.1000358 PMID: 19390598

13. Schnell S. Ten Simple Rules for a Computational Biologist's Laboratory Notebook. PLoS Comput Biol. 2015; 11(9):e1004385. https://doi.org/10.1371/journal.pcbi.1004385 PMID: 26356732

14. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comput Biol. 2013; 9(10):e1003285. https://doi.org/10.1371/journal.pcbi.1003285 PMID: 24204232

15. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, da Veiga Leprevost F, et al. Ten Simple Rules for Taking Advantage of Git and GitHub. PLoS Comput Biol. 2016; 12(7):e1004947. https://doi.org/10.1371/journal.pcbi.1004947 PMID: 27415786

16. Haddock SHD, Dunn CW. Practical computing for biologists: Sinauer Associates Sunderland, MA; 2011.

17. Markowetz F. All biology is computational biology. PLoS Biol. 2017; 15(3):e2002050. https://doi.org/10.1371/journal.pbio.2002050 PMID: 28278152

18. Bergman C. An Assembly of Fragments [Internet]. [cited 2017]. Available from: https://caseybergman.wordpress.com/2012/07/31/top-n-reasons-to-do-a-ph-d-or-post-doc-in-bioinformaticscomputational-biology/. Accessed on 5 April 2017.

19. Kwok R. Nature: Careers [Internet]: Nature Publishing Group. 2013. [cited 2017].

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for making research software more robust

**Morgan Taschuk[1☯*], Greg Wilson[2☯]**

**1** Genome Sequence Informatics, Ontario Institute for Cancer Research, Toronto, Ontario, Canada,
**2** Software Carpentry Foundation, Austin, Texas, United States of America

☯ These authors contributed equally to this work.
* morgan.taschuk@oicr.on.ca

## Abstract

Software produced for research, published and otherwise, suffers from a number of common problems that make it difficult or impossible to run outside the original institution or even off the primary developer's computer. We present ten simple rules to make such software robust enough to be run by anyone, anywhere, and thereby delight your users and collaborators.



## Author summary

Many researchers have found out the hard way that there's a world of difference between "works for me on my machine" and "works for other people on theirs." Many common challenges can be avoided by following a few simple rules; doing so not only improves reproducibility but can accelerate research.

OPEN ACCESS

**Citation:** Taschuk M, Wilson G (2017) Ten simple rules for making research software more robust. PLoS Comput Biol 13(4): e1005412. https://doi.org/10.1371/journal.pcbi.1005412

**Published:** April 13, 2017



**Funding:** This work was partially funded by the Ontario Institute for Cancer Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Scientific software is typically developed and used by a single person, usually a graduate student or postdoc [1]. It may produce the intended results in their hands, but what happens when someone else wants to run it? Everyone with a few years of experience feels a bit nervous when told to use another person's code to analyze their data: it will often be undocumented, work in unexpected ways (if it works at all), rely on nonexistent paths or resources, be tuned for a single dataset, or simply be an older version than was used in published papers. The potential new user is then faced with two unpalatable options: hack the existing code to make it work or start over.

Being unable to replicate results is so common that one publication refers to it as "a rite of passage" [2]. The root cause of this problem is that most research software is essentially a prototype, and therefore is not robust. The lack of robustness in published, distributed software leads to duplicated efforts with little practical benefit, which slows the pace of research [3, 4]. Bioinformatics software repositories [5, 6] catalogue dozens to hundreds of tools that perform similar tasks: for example, in 2016, the Bioinformatics Links Directory included 84 different multiple sequence aligners, 141 tools to analyze transcript expression, and 182 pathway and interaction resources. Some of these tools are legitimate efforts to improve the state-of-the-art, but often, they are difficult to install and run [7, 8] and are effectively abandoned after publication [9].

This problem is not unique to bioinformatics or even to computing [2]. Best practices in software engineering specifically aim to increase software robustness. However, most bioinformaticians learn what they know about software development on the job or otherwise informally [1, 10]. Existing training programs and initiatives rarely have the time to cover software engineering in depth, especially since the field is so broad and developing so rapidly [4, 10]. In addition, making software robust is not directly rewarded in science, and funding is difficult to come by [1]. Some proposed solutions to this problem include restructuring educational programs, hiring dedicated software engineers [4, 11], partnering with private sector or grassroots organizations [1, 5], or using specific technical tools like containerization or cloud computing [12, 13]. Each of these requires time and, in some cases, institutional change.

The good news is you don't need to be a professionally trained programmer to write robust software. In fact, some of the best, most reliable pieces of software in many scientific communities are written by researchers [3, 11] who have adopted strong software engineering approaches, have high standards of reproducibility, use good testing practices, and foster strong user bases through constantly evolving, clearly documented, useful, and useable software. In the bioinformatics community, Bioconductor and Galaxy follow this path [12, 14]. Not all scientific software needs to be robust [15], but if you publish a paper about your software, it should, at minimum, satisfy these rules.

So what is "robust" software? We implied above that it is software that works for people other than the original author and on machines other than its creator's. More specifically, we mean that:

- it can be installed on more than one computer with relative ease,
- it works consistently as advertised, and
- it can be integrated with other tools.

Our rules are generic and can be applied to all languages, libraries, packages, documentation styles, and operating systems for both closed-source and open-source software. They are also necessary steps toward making computational research replicable and reproducible: after all, if your tools and libraries cannot be run by others, they cannot be used to verify your results or as a stepping stone for future work [16].

## Rule 1: Use version control

Version control is essential to sustainable software development [17, 18]. In particular, developers will struggle to understand what they have actually built, what it actually does, and what they have actually released without some mechanical way to keep track of changes. They should therefore put everything into version control as soon as it is created, including programs, original field observations, and the source files for papers. Files that can be regenerated as needed, such as the binaries for compiled programs or intermediate files generated during data analysis, should not be versioned; instead, it is often more sensible to use an archiving system for them and store the metadata describing their contents in version control instead [19].

If you are new to version control, it is simplest to treat it as "a better Dropbox" (or, if you are of a certain age, a better FTP) and to use it simply to synchronize files between multiple developers and machines [20]. Once you are comfortable working that way, you should use a feature branch workflow: designate one parallel copy (or "branch") of the repository as the master, and create a new branch from it each time you want to fix a bug or add a new feature.

This allows work on independent changes to proceed in isolation; once the work has been completed and tested, it can be merged into the master branch for release.

## Rule 2: Document your code and usage

How to write high-quality documentation has been described elsewhere [21], and so here, we only cover two minimal types: the README and usage. The README is usually available even before the software is installed, exists to get a new user started, and points them towards more help. Usage is a terse, informative command-line help message that guides the user in the correct use of the software.

Numerous guidelines exist on how to write a good README file [22, 23]. At a minimum, your README should:

1. Explain what the software does. There's nothing more frustrating than downloading and installing something only to find out that it doesn't do what you thought it did.
2. List required dependencies. We address dependencies in more detail in Rule 5.
3. Provide compilation or installation instructions.
4. List all input and output files, even those considered self-explanatory. Link to specifications for standard formats and list the required fields and acceptable values in other files. If there is no rigorous definition for a format, explain its parts as clearly as possible in plain English.
5. List a few example commands to get a user started quickly.
6. State attributions and licensing. Attributions are how you credit your contributors; licenses dictate how others may use and need to credit your work.

The program should also print usage information when launching from the command line. Usage provides the first line of help for both new and experienced users. Terseness is important: usage that extends for multiple screens is difficult to read or refer to on the fly.

Almost all command-line applications use a combination of POSIX [24] and GNU [23] standards for usage. More standard command-line behaviours are detailed in [8]. Your software's usage should:

1. Describe the syntax for running the program, including the name of the program, the relative location of optional and required flags, other arguments, and values for execution.
2. Give a short description to remind users of the software's primary function.
3. List the most commonly used arguments, a description of each, and the default values.
4. State where to find more information.

Usage should be printed to standard output so that it can be combined with other bash utilities like `grep`, and it should finish with an appropiate exit code.

Documentation beyond the README and usage is up to the developer's discretion. We think it is very important for developers to document their work, but our experience is that people are unlikely do it during normal development. However, it is worth noting that software that is widely used and contributed to has and enforces the need for good documentation [14].

## Rule 3: Make common operations easy to control

Being able to change parameters on the fly to determine if and how they change the results is important as your software gains more users since it facilitates exploratory analysis and parameter sweeping. Programs should therefore allow the most commonly changed parameters to be configured from the command line.

Users will want to change some values more often than others. Since parameters are software-specific, the appropriate "tunable" ones cannot be detailed here, but a short list includes input and reference files and directories, output files and directories, filtering parameters, random number generation seeds, and alternatives such as compressing results, using a variant algorithm, or verbose output.

Check that all input values are in a reasonable range at startup. Few things are as annoying as having a program announce after running for two hours that it isn't going to save its results because the requested directory doesn't exist.

To make programs even easier to use, choose reasonable defaults when they exist and set no defaults at all when there aren't any reasonable ones. You can set reasonable default values as long as any command line arguments override those values.

Changeable values should never be hard-coded: if users have to edit your software in order to run it, you have done something wrong. Changeable but infrequently changed values should therefore be stored in configuration files. These can be in a standard location, e.g., `.packagerc` in the user's home directory, or provided on the command line as an additional argument. Configuration files are often created during installation to set up such things as server names, network drives, and other defaults for your lab or institution.

## Rule 4: Version your releases

Software evolves over time, with developers adding or removing features as need dictates. Making official releases stamps a particular set of features with a project-specific identifier so that version can be retrieved for later use. For example, if a paper is published, the software should be released at the same time so that the results can be reproduced.

Most software has a version number composed of a decimal number that increments as new versions are released. There are many different ways to construct and interpret this number, but most importantly for us, a particular software version run with the same parameters should give identical results no matter when it's run. Results include both correct output as well as any errors. Increment your version number every time you release your software to other people.

Semantic versioning [25] is one of the most common types of versioning for open-source software. Version numbers take the form of "MAJOR.MINOR[.PATCH]," e.g., 0.2.6. Changes in the major version number herald significant changes in the software that are not backwards compatible, such as changing or removing features or altering the primary functions of the software. Increasing the minor version represents incremental improvements in the software, like adding new features. Following the minor version number can be an arbitrary number of project-specific identifiers, including patches, builds, and qualifiers. Common qualifiers include `alpha`, `beta`, and `SNAPSHOT`, for applications that are not yet stable or released, and `-RC` for release candidates prior to an official release.

The version of your software should be easily available by supplying `--version` or `-v` on the command line. This command should print the software name and version number, and it should also be included in all of the program's output, particularly debugging traces. If someone needs help, it's important that they be able to tell whoever's helping them which version of the software they're using.

While new releases may make a program better in general, they can simultaneously create work for someone who integrated the old version into their own workflow a year or two ago and won't see any benefits from upgrading. A program's authors should therefore ensure that old released versions continue to be available. A number of mechanisms exist for controlled release that range from adding an appropriate commit message or tag to version control [20] to official releases alongside code on Bitbucket or GitHub to depositing into a repository like apt, yum, homebrew, CPAN, etc. Choose the method that best suits the number and expertise of users you anticipate.

## Rule 5: Reuse software (within reason)

In the spirit of code reuse and interoperability, developers often want to reuse software written by others. With a few lines, a call is made out to another library or program and the results are incorporated into the primary script. Using popular projects reduces the amount of code that needs to be maintained and leverages the work done by the other software.

Unfortunately, reusing software (whether software libraries or separate executables) introduces dependencies, which can bring their own special pain. The interface between two software packages can be a source of considerable frustration: all too often, support requests descend into debugging errors produced by the other project due to incompatible libraries, versions, or operating systems [16]. Even introducing libraries in the same programming language can rely on software installed in the environment, and the problem becomes much more difficult when relying on executables or even on web services.

Despite these problems, software developers in research should reuse existing software provided a few guidelines are adhered to.

First, make sure that you really need the auxiliary program. If you are executing GNU sort instead of figuring out how to sort lists in Python, it may not be worth the pain of integration. Reuse software that offers some measurable improvement to your project.

Second, if launching an executable, ensure the appropriate software and version is available. Either allow the user to configure the exact path to the package, distribute the program with the dependent software, or download it during installation using your package manager. If the executable requires internet access, check for that early in execution.

Third, ensure that reused software is robust. Relying on erratic third party libraries or software is a recipe for tears. Prefer software that follows good software development practices, is open for support questions, and is available from a stable location or repository using your package manager.

Exercise caution, especially when transitioning across languages or using separate executables, as they tend to be especially sensitive to operating systems, environments, and locales.

## Rule 6: Rely on build tools and package managers for installation

To compile code, deploy applications, and automate other tasks, programmers routinely use build tools like Make, Rake, Maven, Ant, or MS Build. These tools can also be used to manage runtime environments, i.e., to check that the right versions of required packages are installed and install or upgrade them if they are not. As mentioned in Rule 5, a package manager can mitigate some of the difficulties in software reuse.

The same tools can and should be used to manage runtime environments on users' machines as well. Accordingly, developers should document all dependencies in a machine-readable form. Package managers like apt and yum are available on most Unix-like systems, and application package managers exist for specific languages like Python (pip), Java (Maven/Gradle), and Ruby (RubyGems). These package managers can be used together with the build utility to ensure that dependencies are available at compile/run time.

For example, it is common for Python projects to include a file called `requirements.txt` that lists the names of required libraries, along with version ranges:

```
requests>=2.0
pygithub>=1.26,<=1.27
python-social-auth>=0.2.19,<0.3
```

This file can be read by the pip package manager, which can check that the required software is available and install it if it is not. Whatever is used, developers should always install dependencies using their dependency description, especially on their personal machines, so that they're sure it works.

Conversely, developers should avoid depending on scripts and tools which are not available as packages. In many cases, a program's author may not realize that some tool was built locally and doesn't exist elsewhere. At present, the only sure way to discover such unknown dependencies is to install on a system administered by someone else and see what breaks. As use of virtualization containers becomes more widespread, software installation can also be tested on a virtual machine or container system like Docker.

## Rule 7: Do not require root or other special privileges to install or run

Root (also known as "superuser" or "admin") is a special account on a computer that has (among other things) the power to modify or delete system files and user accounts. Conversely, files and directories owned by root usually cannot be modifed by normal users.

Installing or running a program with root privileges is often convenient, since doing so automatically bypasses all those pesky safety checks that might otherwise get in the user's way. However, those checks are there for a reason: scientific software packages may not intentionally be malware, but one small bug or over-eager file-matching expression can certainly make them behave as if they were. Outside of very unusual circumstances, packages should not require root privileges to set up or use.

Another reason for this rule is that users may want to try out a new package before installing it systemwide on a cluster. Requiring root privileges will frustrate such efforts and thereby reduce uptake of the package. Requiring, as Apache Tomcat does, that software be installed under its own user account—i.e., that `packagename` be made a user and all of the package's software be installed in that pseudo-user's space—is similarly limiting, and makes side-by-side installation of multiple versions of the package more difficult.

Developers should therefore allow packages to be installed in an arbitrary location, e.g., under a user's home directory in `~/packagename`, or in directories with standard names like `bin`, `lib`, and `man` under a chosen directory. If the first option is chosen, the user may need to modify his or her search path to include the package's executables and libraries, but this can (more or less) be automated and is much less risky than setting things up as root.

Testing the ability to install software has traditionally been regarded as difficult, since it necessarily alters the machine on which the test is conducted. Lightweight virtualization containers like Docker make this much easier as well, or you can simply ask another person to try and build your software before releasing it.

## Rule 8: Eliminate hard-coded paths

It's easy to write software that reads input from a file called `mydata.csv`, but it's also very limiting. If a colleague asks you to process his or her data, you must either overwrite your data file (which is risky) or edit your code to read `otherdata.csv` (which is also risky, because there's every likelihood you'll forget to change the filename back or will change three uses of the filename but not a fourth).

Hard-coding file paths in a program also makes the software harder to run in other environments. If your package is installed on a cluster, for example, the user's data will almost certainly not be in the same directory as the software, and the folder `C:\users\yourname\` will probably not even exist.

For these reasons, users should be able to set the names and locations of input and output files as command-line parameters. This rule applies to reference datasets as well as the user's own data: if a user wants to try a new gene identification algorithm using a different set of genes as a training set, he or she should not have to edit the software to do so. A corollary to this rule is to not require users to navigate to a particular directory to do their work, since "where I have to be" is just another hard-coded path.

In order to save typing, it is often convenient to allow users to specify an input or output directory, and then require that there be files with particular names in that directory. This practice is an example of "convention over configuration," a principle used by software frameworks such as WordPress and Ruby on Rails that often strikes a good balance between adaptability and consistency.

## Rule 9: Include a small test set that can be run to ensure the software is actually working

Every package should come with a set of tests for users to run after installation. Its purpose is not only to check that the software is working correctly (although that is extremely helpful) but also to ensure that it works at all. This test script can also serve as a working example of how to run the software.

In order to be useful, make the tests easy to find and run. Many build systems will also run unit tests if provided them at compile time. For users, or if the build system is not amenable to testing, provide a working script in the project's root directory named `runtests.sh` or something equally obvious. This lets new users build their analysis from a working script. For example, with its distribution, the graph-based sequence aligner HISAT2 includes a full set of very small files, and a "Getting Started with HISAT2" section in its manual that leads you through the entire data lifecycle [26].

Equally important is to make the test script's output easy to interpret. Screens full of correlation coefficients do not qualify: instead, the script's output should be simple to understand for nonexperts, such as one line per test, with the test's name and its pass/fail status, followed by a single summary line saying how many tests were run and how many passed or failed. If many or all tests fail because of missing dependencies, that fact should be displayed once, clearly, rather than once per test, so that users have a clear idea of what they need to fix and how much work it's likely to take.

Research has shown that the ease with which people can start making contributions is a strong predictor of whether they will or not [27]. By making it simpler for outsiders to contribute, a test suite of any kind also makes it more likely that they will, and software with collaborators stands a better chance of surviving in the busy field of scientific software.

## Rule 10: Produce identical results when given identical inputs

The usage message tells users what the program could do. It is equally important for the program to tell users what it actually did. Accordingly, when the program starts, it should echo all parameters and software versions to standard out or a log file alongside the results to increase the reproducibility of that step.

Given a set of parameters and a dataset, a particular version of a program should produce the same results every time it is run to aid testing, debugging, and reproducibility. Even minor changes to code can cause minor changes in output because of floating-point issues, which means that getting exactly the same output for the same input and parameters probably won't work during development, but it should still be a goal for people who have deployed a specific version.

Many applications rely on randomized algorithms to improve performance or runtimes. As a consequence, results can change between runs, even when provided with the same data and parameters. By its nature, this randomness renders strict reproducibility (and, therefore, debugging) more difficult. If even the small test set (#9) produces different results for each run, new users may not be able to tell whether or not the software is working properly. When comparing results between versions or after changing parameters, even small differences can confuse or muddy the comparison. And especially when producing results for publications, grants, or diagnoses, any analysis should be absolutely reproducible.

Given the size of biological data, it is unreasonable to suggest that random algorithms be removed. However, most programs use a pseudo-random number generator, which uses a starting seed and an equation to approximate random numbers. Setting the seed to a consistent value can remove randomness between runs. Allow the user to optionally provide the random seed as an input parameter, thus rendering the program deterministic for those cases where it matters. If the seed is set internally (e.g., using clock time), echo it to the output for reuse later. If setting the seed is not possible, make sure the acceptable tolerance is known and detailed in documentation and in the tests.

## Conclusion

There has been extended discussion over the past few years of the sustainability of research software, but this question is meaningless in isolation: any piece of software can be sustained if its users are willing to put in enough effort. The real equation is the ratio between the skill and effort available and the ease with which software can be installed, understood, used, maintained, and extended. Following the ten rules we outline here reduces the denominator and thereby enables researchers to build on each other's work more easily.

That said, not every coding effort needs to be engineered to last. Code that is used once to answer a specific question related to a specific dataset doesn't require comprehensive documentation or flexible configuration, and the only sensible way to test it may well be to run it on the dataset in question. Exploratory analysis is an iterative process that is developed quickly and revised often [4, 11]. However, if a script is dusted off and run three or four times for slightly different purposes, is crucial to a publication or a lab, or is being passed on to someone else, it may be time to make your software more robust.

## Supporting information

**S1 Checklist. Robust software checklist.** A checklist summarizing these ten simple rules to apply to your own software.
(PDF)

## Acknowledgments

The authors thank the reviewers who provided feedback on the manuscript before and after submission. The list is available on our Github page: http://github.com/oicr-gsi/robust-paper.

## References


1. Prins P, de Ligt J, Tarasov A, Jansen RC, Cuppen E, Bourne PE. Toward effective software solutions for big biology. Nature Biotechnology. 2015; 33(7):686–687. https://doi.org/10.1038/nbt.3240 PMID: 26154002
2. Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016; 533(7604):452–454. https://doi.org/10.1038/533452a PMID: 27225100
3. Prabhu P, Jablin TB, Raman A, Zhang Y, Huang J, Kim H, et al. A Survey of the Practice of Computational Science. In: State of the Practice Reports. SC'11. New York, NY, USA: ACM; 2011. p. 19:1–19:12.
4. Lawlor B, Walsh P. Engineering bioinformatics: building reliability, performance and productivity into bioinformatics software. Bioengineered. 2015; 6(4). https://doi.org/10.1080/21655979.2015.1050162 PMID: 25996054
5. Ison J, Rapacki K, Ménager H, Kalaš M, Rydza E, Chmura P, et al. Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Research. 2016; 44(D1):D38–D47. https://doi.org/10.1093/nar/gkv1116 PMID: 26538599
6. Brazas MD, Yim D, Yeung W, Ouellette BFF. A decade of web server updates at the bioinformatics links directory: 2003–2012. Nucleic Acids Research. 2012; https://doi.org/10.1093/nar/gks632 PMID: 22700703
7. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, et al. The Bioperl Toolkit: Perl Modules for the Life Sciences. Genome Research. 2002; 12(10):1611–1618. https://doi.org/10.1101/gr.361602 PMID: 12368254
8. Seemann T. Ten recommendations for creating usable bioinformatics command line software. GigaScience. 2013; 2(1):15. https://doi.org/10.1186/2047-217X-2-15 PMID: 24225083
9. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nature ReviewsGenetics. 2012; 13(9):667–72.
10. Atwood TK, Bongcam-Rudloff E, Brazas ME, Corpas M, Gaudet P, Lewitter F, et al. GOBLET: The Global Organisation for Bioinformatics Learning, Education and Training. PLoS Computational Biology. 2015; https://doi.org/10.1371/journal.pcbi.1004143 PMID: 25856076
11. Sanders R, Kelly D. Dealing with Risk in Scientific Software Development. Software, IEEE. 2008; 25 (4):21–28. https://doi.org/10.1109/MS.2008.84
12. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Research. 2016; 44(W1):W3–W10. https://doi.org/10.1093/nar/gkw343 PMID: 27137889
13. Howe B. Virtual Appliances, Cloud Computing, and Reproducible Research. Computing in Science Engineering. 2012; 14(4):36–41. https://doi.org/10.1109/MCSE.2012.62
14. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology. 2004; 5(10):1–16. https://doi.org/10.1186/gb-2004-5-10-r80
15. Varoquaux G. Software for reproducible science: let's not have a misunderstanding; 2015. http://gael-varoquaux.info/programming/software-for-reproducible-science-lets-not-have-a-misunderstanding.html.
16. Brown CT. Replication, reproduction, and remixing in research software; 2013. http://ivory.idyll.org/blog/research-software-reuse.html.
17. Wilson G, Aruliah DA, Brown CT, Hong NPC, Davis M, Guy RT, et al. Best Practices for Scientific Computing. PLoS Biology. 2014; 12(1):e1001745. https://doi.org/10.1371/journal.pbio.1001745 PMID: 24415924
18. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good Enough Practices in Scientific Computing. arxivorg. 2016;abs/1609.00037.
19. Noble WS. A Quick Guide to Organizing Computational Biology Projects. PLoS Computational Biology. 2009; 5(7). https://doi.org/10.1371/journal.pcbi.1000424 PMID: 19649301
20. Blischak JD, Davenport ER, Wilson G. A Quick Introduction to Version Control with Git and GitHub. PLOS Computational Biology. 2016; 12(1):1–18. https://doi.org/10.1371/journal.pcbi.1004668
21. Karimzadeh M, Hoffman MM. Top considerations for creating bioinformatics software documentation. Briefings in Bioinformatics. in press 2017; https://doi.org/10.1093/bib/bbw134 PMID: 28088754
22. Johnson M. Building a Better ReadMe. Technical Communication. 1997; 44(1):28–36.
23. Free Software Foundation. GNU Coding Standards; 2016. https://www.gnu.org/prep/standards/standards.html.

24. The IEEE, The Open Group. The Open Group Base Specifications Issue 7 IEEE Std 1003.1-2008. 12. Utility Conventions; 2016. http://pubs.opengroup.org/onlinepubs/9699919799/basedefs/V1_chap12.html.
25. Preston-Werner T. Semantic Versioning 2.0.0; 2016. http://semver.org/spec/v2.0.0.html.
26. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature Protocols. 2016; 11(9):1650–1667. https://doi.org/10.1038/nprot.2016.095 PMID: 27560171
27. Steinmacher I, Silva MAG, Gerosa MA, Redmiles DF. A systematic literature review on the barriers faced by newcomers to open source software projects. Information and Software Technology. 2015; 59:67–85. http://dx.doi.org/10.1016/j.infsof.2014.11.001.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Developing Usable Software in Computational Biology

**Markus List**[1☯*], **Peter Ebert**[1,2☯*], **Felipe Albrecht**[1,2]

**1** Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany, **2** Graduate School of Computer Science, Saarland Informatics Campus, Saarbrücken, Germany

☯ These authors contributed equally to this work.
* markus.list@mpi-inf.mpg.de (ML); pebert@mpi-inf.mpg.de (PE)

CrossMark
click for updates

OPEN ACCESS

**Citation:** List M, Ebert P, Albrecht F (2017) Ten Simple Rules for Developing Usable Software in Computational Biology. PLoS Comput Biol 13(1): e1005265. doi:10.1371/journal.pcbi.1005265

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** January 5, 2017



**Funding:** FA and PE acknowledge the support of the German Federal Ministry of Education and Research grant no. 01KU1216A (DEEP project). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

The rise of high-throughput technologies in molecular biology has led to a massive amount of publicly available data. While computational method development has been a cornerstone of biomedical research for decades, the rapid technological progress in the wet lab makes it difficult for software development to keep pace. Wet lab scientists rely heavily on computational methods, especially since more research is now performed in silico. However, suitable tools do not always exist, and not everyone has the skills to write complex software. Computational biologists are required to close this gap, but they often lack formal training in software engineering. To alleviate this, several related challenges have been previously addressed in the Ten Simple Rules series, including reproducibility [1], effectiveness [2], and open-source development of software [3, 4].

Here, we want to shed light on issues concerning software usability. Usability is commonly defined as "a measure of interface quality that refers to the effectiveness, efficiency, and satisfaction with which users can perform tasks with a tool" [5]. Considering the subjective nature of this topic, a broad consensus may be hard to achieve. Nevertheless, good usability is imperative for achieving wide acceptance of a software tool in the community. In many cases, academic software starts out as a prototype that solves one specific task and is not geared for a larger user group. As soon as the developer realizes that the complexity of the problems solved by the software could make it widely applicable, the software will grow to meet the new demands. At least by this point, if not sooner, usability should become a priority. Unfortunately, efforts in scientific software development are constrained by limited funding, time, and rapid turnover of group members. As a result, scientific software is often poorly documented, non-intuitive, non-robust with regards to input data and parameters, and hard to install. For many use cases, there is a plethora of tools that appear very similar and make it difficult for the user to select the one that best fits their needs. Not surprisingly, a substantial fraction of these tools are probably abandonware; i.e., these are no longer actively developed or supported in spite of their potential value to the scientific community.

To our knowledge, software development as part of scientific research is usually carried out by individuals or small teams with no more than two or three members. Hence, the responsibility of designing, implementing, testing, and documenting the code rests on few shoulders. Additionally, there is pressure to produce publishable results or, at least, to contribute analysis work to ongoing projects. Consequently, academic software is typically released as a prototype. We acknowledge that such a tool cannot adhere to and should not be judged by the standards that we take for granted for production grade software. However, widespread use of a tool is typically in the interest of a researcher. To this end, we propose ten simple rules that, in our experience, have a considerable impact on improving usability of scientific software.

## Rule 1: Identify the Missing Pieces

Unless you are a pioneer, and few of us are, the problem you are working on is likely addressed by existing tools. As a professional, you are aware of this software but may consider it cumbersome, non-functional, or otherwise unacceptable for your demands. Make sure that your judgment is shared by a substantial fraction of the prospective users before you start developing a new tool. Usable software should offer the features needed and behave as expected by the community. Moreover, a new tool needs to provide substantial novelty over existing solutions. For this purpose, list the requirements on the software and create a comparison table to set the new tool against existing solutions. This allows you to carve out the selling points of your tool in a systematic fashion.

## Rule 2: Collect Feedback from Prospective Users

Software can be regarded as providing the interface between wet lab science and data analysis. A lack of communication between both sides will lead to misunderstandings that need to be rectified by substantially changing the code base in a late phase of the project. Avoid this pitfall by exposing potential users to a prototype. Discussions on data formats or on the design of the user interface will reveal unforeseen challenges and help to determine if a tool is sufficiently intuitive [6]. To plan your progress, keep a record of suggested improvements and existing issues.

## Rule 3: Be Ready for Data Growth

First estimate the expected data growth in your field and then design your software accordingly. To this end, consider parallelization and make sure your tool can be integrated seamlessly in workflow management systems (e.g., GALAXY [7] and Taverna [8]), pipeline frameworks (e.g., Ruffus [9] and SnakeMake [10]), or a cluster framework (e.g., Hadoop, http://hadoop.apache.org/). Moreover, make sure that the user interface can scale to growing data volumes. For example, consider that the visualizations should still be comprehensible for larger datasets, e.g., by displaying only parts of the data or through aggregation of results.

## Rule 4: Use Standard Data Formats for Input and Output

As an expert in your research domain, you know the established data standards and related programming libraries for reading and writing commonly used data formats. Make sure that your tool's output follows standard specifications to the letter, but be as lenient as possible when users provide non-standard input. Tools that follow this rule are more likely to become successful. If you are working in an emerging field with no prevalent model for data exchange, provide data in a structured text file (e.g., tab-separated tables, XML/XSD, or JSON) and aim for self-documenting output by including header lines and data type descriptions. In this case, document how users can derive suitable input data for your tool.

## Rule 5: Expose Only Mandatory Parameters

Exposing all (possible) parameters to a user can be confusing and carries the risk of nonsensical parameters settings. When possible, users will thus rely on default parameters. The same applies to benchmark studies comparing your tool against the state-of-the-art competitors.

This has three important implications: (i) expose only a small set of parameters by default whose effects on results can be easily understood by any user, (ii) offer advanced parameters only in an expert section and describe them thoroughly in the documentation, and (iii) choose conservatively (and if possible, justify) the default values for parameters such that the tool can operate in a wide range of scenarios and within reasonable run time.

## Rule 6: Expect Users to Make Mistakes

You should never assume that your tool is self-explanatory, that requirements concerning the input data are obvious, or that the user will immediately grasp all details of the problem at hand. Ideally, your tool supports the user in using it appropriately, e.g., by checking that data remain inside required ranges or that identifiers are unique, and provides descriptive error messages in case of unexpected values. If performance penalties due to such checks are a real concern (which should be tested), make the checks optional and enabled by default. Finally, allow users to stop ongoing operations in case they realize they made a mistake.

## Rule 7: Provide Logging Information

Two types of logs improve usability and also support the user in making their research more reproducible. Configuration logs keep track of basic information, such as the time stamp of the analysis, the version of your tool and of third-party libraries, as well as the parameter settings and input data. Archiving this information is particularly important in long-running research projects in order to trace irregularities in the results at any later point in time [1]. Technical logs, on the other hand, contain progress messages that help users to pinpoint errors in the execution flow and allow clear communication of these issues to the developer. As much as possible, avoid exposing potentially sensitive user information in the logs.

## Rule 8: Get Users Started Quickly

Complex setup routines introduce dependency [11] or configuration debt [12]; i.e., the user has to spend substantial time installing software and learning about the execution parameters of a tool. These raise the bar for unhindered exploration of software features. Such issues can be solved by implementing a web application (if feasible with respect to resource demands), by providing a standalone executable, or by providing a system-specific software package. Alternatively, issues of a program's dependence on third-party libraries can be avoided by encapsulating your tool in a virtual machine image or, e.g., a Docker container (https://docker.com). Finally, it is imperative to provide demo data that enable users to immediately interact with the software. A successful test run proves to the user that your software works as expected and will be essential if you want your tool to be published.

## Rule 9: Offer Tutorial Material

Researchers can seldom afford the time to thoroughly read complex user manuals. They will thus appreciate a number of clearly written code examples, illustrations, or video screen casts to get started. Most importantly, documented use cases enable users to quickly assess if your tool is suited for the problem at hand and allow fast learning by doing. Keep in mind that these materials have to be updated together with your tool.

## Rule 10: Consider the Future of Your Tool

For long-term availability of your software, use suitable repositories such as github (https://github.com) or bitbucket (https://bitbucket.com) throughout the development process.

Explicitly state under which software license you release your code for third parties (see https://opensource.org/licenses). Without such a license, using your software might be prohibitive for many organizations or companies. More importantly, keeping your code in a public repository will also allow you to engage with the users through issue tracking (e.g., bugs, suggestions). After releasing your tool, expect support requests and take them seriously. See them as an opportunity to continuously improve the usability of your tool.

## Conclusions

Usability is an important topic in software design, and we would like to provide a few starting points for further reading [13–18]. In the above ten simple rules, we highlight that software should not only be scientifically sound but also be perceived as usable for widespread and effective application. To these ends, developers should also be the first to apply their tool, to reveal usability issues as early as possible. However, effort is required from both users and developers to further improve a tool. Even engaging with only a few users (Rule 2) is likely to have a large impact on usability, since, as Jakob Nielsen put it, "Zero users give zero insights" [19].

## Acknowledgments

We would like to thank Thomas Lengauer, Nico Pfeifer, and Fabian Müller for their critical reading of the manuscript and insightful comments.

## References


1. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comp Biol. 2013;. doi: 10.1371/journal.pcbi.1003285 PMID: 24204232
2. Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, Dalchau N, et al. Ten simple rules for effective computational research. PLoS Comput Biol. 2014; 10(3):e1003506. doi: 10.1371/journal.pcbi.1003506 PMID: 24675742
3. Prlić A, Procter JB. Ten simple rules for the open development of scientific software. PLoS Comput Biol. 2012; 8(12):e1002802. doi: 10.1371/journal.pcbi.1002802 PMID: 23236269
4. Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost FdV, et al. Ten Simple Rules for Taking Advantage of Git and GitHub. PLOS Computational Biology. 2016 jul; 12(7):e1004947. doi: 10.1371/journal.pcbi.1004947 PMID: 27415786
5. Dillon A. Encyclopedia of Human Factors and Ergonomic. Ed. London: Taylor and Francis, ch. User Acceptance of Information Technology; 2001.
6. Thielsch MT, Engel R, Hirschfeld G. Expected usability is not a valid indicator of experienced usability. PeerJ Computer Science. 2015; 1:e19. doi: 10.7717/peerj-cs.19
7. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy: a platform for interactive large-scale genome analysis. Genome research. 2005; 15(10):1451–1455. doi: 10.1101/gr.4086505 PMID: 16169926
8. Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S, et al. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic acids research. 2013;p. gkt328. doi: 10.1093/nar/gkt328 PMID: 23640334
9. Goodstadt L. Ruffus: a lightweight Python library for computational pipelines. Bioinformatics. 2010; 26(21):2778–2779. doi: 10.1093/bioinformatics/btq524 PMID: 20847218
10. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012; 28(19):2520–2522. doi: 10.1093/bioinformatics/bts480 PMID: 22908215
11. Morgenthaler JD, Gridnev M, Sauciuc R, Bhansali S. Searching for build debt: Experiences managing technical debt at Google. In: Proceedings of the Third International Workshop on Managing Technical Debt. IEEE Press; 2012. p. 1–6.
12. Sculley D, Holt G, Golovin D, Davydov E, Phillips T, Ebner D, et al. Machine Learning: The High Interest Credit Card of Technical Debt. In: SE4ML: Software Engineering for Machine Learning (NIPS 2014 Workshop); 2014.

13. Nichols D, Twidale M. The Usability of Open Source Software. First Monday. 2003; 8(1). Available from: http://firstmonday.org/ojs/index.php/fm/article/view/1018.

14. Seffah A, Metzker E. The obstacles and myths of usability and software engineering. Communications of the ACM. 2004; 47(12):71–76. doi: 10.1145/1035134.1035136

15. Macaulay C, Sloan D, Jiang X, Forbes P, Loynton S, Swedlow JR, et al. Usability and User-Centered Design in Scientific Software Development. IEEE Software. 2009 Jan; 26(1):96–102. Copyright—Copyright IEEE Computer Society Jan/Feb 2009; Document feature—; Last updated—2012-07-19; CODEN—IESOEG. Available from: http://search.proquest.com/docview/215838611?accountid=104681. doi: 10.1109/MS.2009.27

16. Sloan D, Macaulay C, Forbes P, Loynton S. User research in a scientific software development project. In: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology. British Computer Society; 2009. p. 423–429.

17. Baxter SM, Day SW, Fetrow JS, Reisinger SJ. Scientific software development is not an oxymoron. PLoS Comput Biol. 2006; 2(9):e87. doi: 10.1371/journal.pcbi.0020087 PMID: 16965174

18. Borchardt JC. Usability in free software;. Last access: 2016-09-30. Available from: http://jancborchardt.net/usability-in-free-software.

19. Nielsen J. Why you only need to test with 5 users;. Last access: 2016-09-30. Available from: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Developing Public Biological Databases

**Mohamed Helmy[1], Alexander Crits-Christoph[2], Gary D. Bader[1]***

**1** The Donnelly Centre, University of Toronto, Toronto, ON, Canada, **2** Department of Biophysics, Johns Hopkins University, Baltimore, MD, United States of America

* gary.bader@utoronto.ca

## Introduction

Biological databases are online libraries that contain structured information about living organisms. These databases are indispensable research tools, as they provide convenient, computable access to prior knowledge that is vital for planning future experiments and for discovering new knowledge through data mining—they help us "stand on the shoulders of giants." Because of their importance to research, the number of public biological databases is increasing. For instance, the number of biological databases published per year in the journal *Nucleic Acid Research* (*NAR*) increased dramatically from only two databases in 1980 to 182 in 2016, with the expectation that this single journal will have published over 2,500 database articles by the end of 2017 [1]. Some of these databases are key, sophisticated, user-friendly, long-term, stable resources, built and maintained by professional teams. However, others have been criticized for being difficult to use or having unclear data quality levels [2,3], and many become obsolete over time [4]. So, if you are considering developing a new database, and especially if you are a student or postdoc, please, for the love of science, follow these ten simple rules for creating and maintaining biological databases (and also a similar set of great rules for scientific web resources [5]).

## Rule 1: Don't reinvent the wheel

Creating a high-quality database is a responsibility that involves strong commitment to accurate data collection and regular content and feature updates, not to mention a substantial time investment. The strongest reason to create a new database is scientific demand for a type of data not easily available in a computable form anywhere else. It is most useful to have all data of a single type in one easy-to-search location, so, ideally, everyone interested in collecting data about a specific topic should collaborate to create one resource, or at least should coordinate efforts to reduce duplication of work (Fig 1). Either way, the data content and software features you create will have the greatest impact if they are original and useful; thus, a comprehensive literature review is a necessary starting point that comes before any actual work (as in any scientific endeavor). A good place to start searching for relevant prior work is the "NAR Online Molecular Biology Database Collection," which, as of January 2016, contains 1,666 biological databases organized into categories [1], collected from the annual *NAR* database issue. *NAR* also publishes an annual web server issue dedicated to web-based software resources [6], and the journal *Database* focuses on biological databases and curation. Several online directories maintain biological database link collections, such as Pathguide (547 databases) [7], The Tools and Data Service Registry (557 databases) [8], The Bioinformatics Links Directory (623 databases) [9], *OMICtools* (1,513 databases) [10], and *MetaBase* (1,802 databases) [11].



**Citation:** Helmy M, Crits-Christoph A, Bader GD (2016) Ten Simple Rules for Developing Public Biological Databases. PLoS Comput Biol 12(11): e1005128. doi:10.1371/journal.pcbi.1005128

**Published:** November 10, 2016



**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

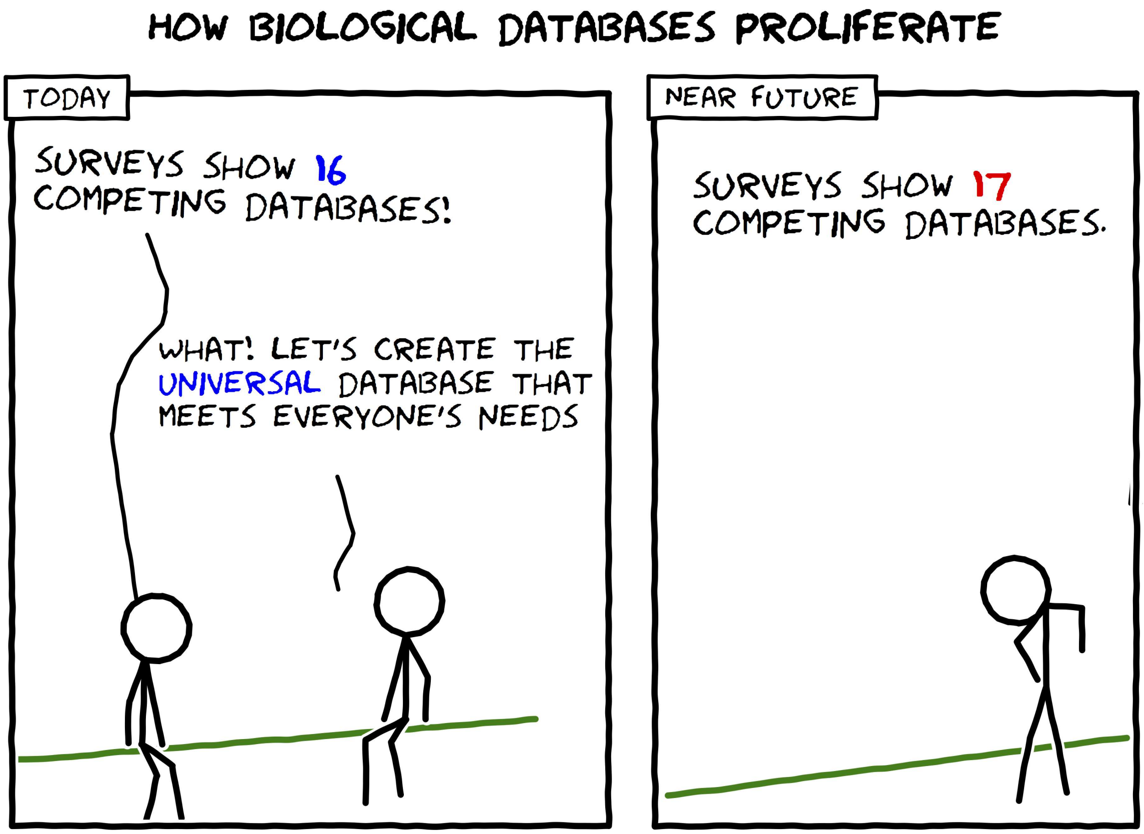



**Fig 1. How biological databases proliferate (adapted from https://xkcd.com/927/ and drawn using Comix I/O [http://cmx.io/]).**

doi:10.1371/journal.pcbi.1005128.g001

## Rule 2: The three most important things in database development are data quality, data quality, and data quality

Some databases collect unique content directly from experiments via authors (e.g., GenBank [12]) or expertly curated from the literature (here, we call these "primary"), and some collect non-unique content from other databases (we call these "secondary" or meta-databases). Primary data could be collected from a community of data generators, such as with GenBank [12], or curated by experts, such as with UniProtKB/Swiss-Prot [13]. Secondary databases, such as InterPro [14], host data collected from other public resources that are often formatted and processed in a uniform way, which can create a more comprehensive and useful data source. Depending on the database type, there are different standards for quality control. Primary databases are responsible for thoroughly checking the quality of all input data, for instance using manual cross-checking and rule-based automated data entry validation [15,16]. It is a good idea to develop a manual describing standard operating procedures for curation to help maintain and communicate quality standards. Secondary databases must ensure to not introduce new errors via their data integration process. Ideally, the integration process will improve data quality via cleaning and normalizing (e.g., standard database identifier use). In either case, following best practices and standards can help improve data quality. A data model standard may require that particular information (e.g., database or gene identifier, an ontology) is used, otherwise intended uses will not be possible. Often, an automatic data validator is available for a standard format to ensure that rules are followed, leading to more consistency and fewer errors. For meta-databases, transforming all incoming data to a standard format eases data integration and can reduce software development time and implementation errors, as well-tested software libraries are available to help read, write, and access the data. For example, the Proteomics Standards Initiative Molecular Interaction (PSI-MI) format captures molecular interactions, such as protein–protein interactions, and maintains a validator, a software library, and other tools to ease working with this data type [17]. Data provenance, information about where the data has come from and when it was updated, is also important to capture and track in order to help users critically evaluate database content quality.

## Rule 3: Know your audience

OK, so there is more to database development than data quality. Defining a target audience is critical for defining database scope and useful data access interfaces (e.g., web interface, application programming interface [API]). Use cases supporting this audience should then be defined, which can then be addressed clearly by what data types will be collected and what queries will be emphasized. Users of biological databases are often of three types: (a) biologists and clinical researchers interested in manual queries via the web, (b) computational biologists interested in batch data download and query systems, and (c) software developers interested in APIs. One study showed that two thirds of users of online biological resources have limited programming experience [5]. These users generally require a web interface that facilitates data access and retrieval. On the other hand, computational biologists and software developers generally need multiple channels for accessing data programmatically or in batch for use and integration into their own analysis workflows and software tools. For instance, NCBI provides the Entrez web interface for manual queries across many databases, a download site for batch data access, and an extensive web service API for programmatic access [18].

## Rule 4: Use modern technology

Most databases are available via an interactive user interface on the web catering to the majority of users. Modern users demand that websites adhere to very high technical and design standards, as is commonplace with websites like Facebook and Google. Websites should have intuitive graphical design and support standard web browsers and operating systems. They also must have a smooth and responsive graphical user interface (GUI) and be secure and robust. Popular technologies such as HTML5, CSS3, and JavaScript now make it straightforward to create such websites. Use of standard front-end technologies often automatically provides support for multiple browsers and operating systems and makes it easy to implement the concept of Responsive Web Design (RWD), which enables a website to adapt its layout to optimally fit different viewing environments, such as desktop, mobile browsers, or touch devices [19]. It is also now possible to build such websites quickly from reusable components. Front-end templates, such as Twitter Bootstrap, help create standardized pages, icons, symbols, graphical components, and fonts. JavaScript libraries (e.g., JQuery) help create GUI components, such as forms and animations, without much effort. Web application frameworks, such as Shiny for R and Django for Python, are collections of ready-made packages and tools that enable developers to automatically generate whole websites, thus greatly speeding up the development process and reducing the need for debugging and testing. Furthermore, several modern technologies, such as node.js, enable writing both the client and server side in one language, JavaScript, and come with thousands of packages providing powerful features.

There has also been a recent diversification in freely available database technologies. If your data content is naturally tabular with a few fields connecting each table with another, it is likely to be relational and can be stored in databases such as MySQL. If content is more naturally organized as free-text or structured documents, networks, key-value pairs, or other non-tabular data, "NoSQL" database technology may be a better choice. Some database systems also make it easy to replicate a database across multiple servers, making indexing and retrieval of large quantities of data faster, because operations can be parallelized. For instance, MongoDB stores structured documents and Apache Lucene indexes free text documents. Using modern technologies in interface and database development can create a speedier user experience and ease the technical maintenance of the system (see Rule 9).

## Rule 5: Put yourself in your user's shoes

A common complaint of researchers using biological databases is that the interfaces to these resources are often difficult to navigate [2,3,20]. Databases with a good user experience make it easy to navigate and find data even by non-experts. Database navigation is facilitated by a graphical user interface that organizes, presents, and visualizes these data in a human accessible manner.

The process of graphical user interface design should be heavily influenced by principles of consistent and appealing graphical design [21], information visualization, and user-specific needs (see Rule 4) [22]. It is useful to employ a process of iterative design, in which feedback from potential users or experienced advisors about the interface is collected before, during, and after the development of the design [23,24]. Interfaces should immediately present the user with the most essential search and browse options and require the least amount of user actions (e.g., keyboard keystrokes, touchscreen taps) to reach the desired information in a speedy and responsive manner. Graphically appealing and color blind–friendly [25] color palettes such as Color Brewer (http://colorbrewer2.org/) are now standard and should be used when choosing a color scheme for both your web interface and data visualization. Interactive data visualization libraries, such as D3.js and BioJS [26,27], are useful sources of visualization methods to display all or selected subsets of large, complex, and heterogeneous biological data sets [28]. For example, jsPhyloSVG [29] draws phylogenetic trees and Cytoscape.js [30] draws networks. Often an artistically talented student can be recruited to help with this, but consider hiring a professional consultant on larger projects.

## Rule 6: Keep search simple and organized

Search options should help users precisely and quickly find what they are looking for. Most users are casual and rely on quick access to data via web-based search functions; thus, it is a good idea to design the default search functionality to meet a small set of the most frequent use cases. Minimally, a keyword search should be available that ideally provides a simple "Google-like" search term input form. It is likely that an advanced query system is both time consuming to implement and will be infrequently used; thus, addressing less frequent use cases can be decided based on resources available or could be left to advanced users to address themselves using available APIs. In all cases, it is useful to include a feature that loads an example query into the search system to help the user quickly figure out the correct input data types and format.

The organization of search results is important for enabling the user to quickly identify relevant results. First, search results should be grouped by the most important data type according to your database content. Second, if your query system accepts multiple simultaneous term searches (e.g., multiple genes in a gene database), group the results by search term. Third, sort the results by relevance, from more relevant to less relevant by following standard data mining approaches, such as term frequency–inverse document frequency (TF-IDF) [31]. These are usually available in standard text indexing systems, like Lucene. Fourth, add refinement or filtering options that enable interactive narrowing of the results, as implemented, for example, in the iRefWeb web interface for protein interaction data [32]. A well-developed refinement option is faceted search [33]. Fifth, show key information first, and then enable the user to ask for more, for instance, by grouping results into tabs or web page sections by data type. Sixth, summarize the results by showing the total number of results, including a breakdown by important categories.

## Rule 7: Give users data where they need it

Most users will likely find data access via a user-friendly, interactive website to be most convenient. However, users working with data in batch will find this to be inconvenient. Instead, they will want to download all the data in easy to process files. It is also useful to release at regular intervals to help users know when they should update and to archive previous data releases to help users reproduce published results based on older versions. Those accessing data from within software will want access to a web service API to program using. Use of your content can be amplified by copying to secondary databases, like GeneCards [34]. Each access channel should include well-written documentation and worked examples showing how to use the system.

## Rule 8: Support open science

Creating a database is a valuable intellectual contribution and a lot of work—you should feel proud to share it. Publish your data model or ontology in a journal and your source code in an open source venue, such as GitHub (http://www.github.com). Sharing source code enables others to contribute, fix the system if the original maintainer loses interest, provide reusable code components for other developers, and provide example code for training purposes. You may even make reusable software a goal of your project, such as with the Generic Model Organism Database (GMOD) for hosting model organism data [35]. Not only does the open access model provide great benefits to the community and is considered a best practice in the field [36], it is increasingly required by journals and funding agencies.

## Rule 9: Tell the world

Building your database and making it available online is unfortunately not enough to get people to use it—it needs to be actively promoted. First, publish an article describing your database in an appropriate journal. We recommend publishing in a journal that is likely to be read by the intended user community, though it may also be useful to publish in a journal devoted to database descriptions, such as the *NAR* database or web service issue, or *Database*: *The Journal of Biological Databases and Curation*. Second, index your website in popular search engines so that your database appears towards the top of the search results when searching related topics. This requires knowledge of search engine optimization techniques such as providing good and unique content, a stable website, and proper indexing of deep content [37]. Third, register your database in specialized online directories that list similar resources, such as *OMICtools* [10]. Fourth, promote your database in scientific conferences and meetings. Fifth, monitor online user groups, such as biostars.org, for potential users and let them know about your resource. Sixth, actively use social media to attract new users and keep them up to date with news about your resource [38]. Using social media can attract many users but requires sustained effort to be effective. If you cannot provide this effort over time, you shouldn't create any social media content, as it will quickly go out of date and reflect badly on your resource (as we discovered in our GeneMANIA project). Finally, it is important to track usage to optimize database usability, promotion activities, and generally to measure how useful the database is to the community.

Accurate tracking is difficult, especially if there are many distributed channels for data access, but web analytic tools, such as Google Analytics (https://analytics.google.com), and monitoring online mentions and citations can help. It is particularly useful to track examples where users have relied on the database to make new discoveries. This can be accomplished by scanning all papers that cite or mention your database and by personally discussing the utility of your database with users at conferences.

## Rule 10: Maintain, update, or retire

Maintaining a database is important for science reproducibility, and many research projects may depend on it being available far into the future. Many journals require a minimum period of at least two years of database maintenance for publication, and funding agencies require sustainable data sharing practices for continued support. Finally, failing to maintain a needed resource that you created can negatively affect your reputation and lead to reduced ability to publish similar work in the future.

Databases aiming for long life are often maintained by major institutions or consortia who employ dedicated staff to maintain content, software, and user support. On the other hand, the majority of databases are developed in small labs and mainly through personnel, such as students, with non-permanent positions [5]. In these cases, the database may become unsustainable because of lack of funding or career changes by interested individuals, which can eventually lead to its disappearance [39].

Fortunately, there are many technologies to help with sustainability on the technology side. First, use professionally managed hosting resources (e.g., cloud, institutional), instead of setting up your own web server. Second, if you must run your own server, use virtualization technology, like Docker, to ease system administration and backup tasks. If the server hardware fails, you can bring the system back quickly on another machine. Third, make your database available for download as a set of files or as an all-in-one virtual machine so that the system can be mirrored by others in case the central server is not available. Virtual machines can be contributed to free online repositories (e.g., https://hub.docker.com/). Fourth, regularly backup (and test restoration) your contents to avoid loss of information due to unfortunate technical problems—this is taken care of if you're on the cloud. Fifth, make your database URL institution-independent (e.g., use.org or.net) to avoid breaking URL changes. Sixth, automatically monitor your system availability and test its main functions (e.g., with https://sensuapp.org/). If your site fails, the system will email you so you can fix it quickly. Seventh, provide means for users to report bugs and request features, such that your database can be improved. GitHub (https://github.com) provides a useful and free issue tracker and is automatically available if you host your source code there. Eighth, choose development technologies that are free and popular, as this will increase the chances that someone else will be available to fix or extend the system if needed. Ninth, if the database is outdated or can no longer be maintained, switch it off, but archive it publicly (e.g., on https://zenodo.org/ or a free virtual machine [VM] repository) so that others can resurrect it if needed [5].

Maintaining content is more challenging, and a significant effort is required to keep the contents of the database up to date, well documented, accurate, and comprehensive [15,16]. For meta-databases that draw content from other sites, content aggregation should be automated at regular intervals from the beginning of the project using tools like snakemake [40]. For high-quality curated databases, there is currently no replacement for the time consuming curation process; however, crowdsourcing is a promising research area that may already be able to help with certain tasks and, in the future, may greatly improve curation efficiency [41]. Sustainability of community-run databases is a hot discussion topic that we expect to generate new solutions for this challenge.

We hope that these ten rules provide a useful checklist and set of pointers to the literature for new database projects. Please visit our dynamic resource for recommended tools, technologies, and libraries at http://baderlab.org/TenRulesResources. Now get out there and collect and share data in a computable form!

## References


1. Rigden DJ, Fernández-Suárez XM, Galperin MY. The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection. Nucleic Acids Res. Oxford University Press; 2016; 44: D1–6. doi: 10.1093/nar/gkv1356 PMID: 26740669
2. Küntzer J, Eggle D, Klostermann S, Burtscher H. Human variation databases. Database (Oxford). 2010; 2010: baq015.
3. Soussi T. Locus-specific databases in cancer: what future in a post-genomic era? The TP53 LSDB paradigm. Hum Mutat. 2014; 35: 643–53. doi: 10.1002/humu.22518 PMID: 24478183
4. Galperin MY, Rigden DJ, Fernández-Suárez XM. The 2015 Nucleic Acids Research Database Issue and molecular biology database collection. Nucleic Acids Res. 2015; 43: D1–5. doi: 10.1093/nar/gku1241 PMID: 25593347
5. Schultheiss SJ. Ten simple rules for providing a scientific Web resource. PLoS Comput Biol. 2011; 7: e1001126. doi: 10.1371/journal.pcbi.1001126 PMID: 21637800
6. Benson G. Editorial: Nucleic Acids Research annual Web Server Issue in 2015. Nucleic Acids Res. 2015; 43: W1–2. doi: 10.1093/nar/gkv581 PMID: 26136473
7. Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list. Nucleic Acids Res. 2006; 34: D504–6. doi: 10.1093/nar/gkj126 PMID: 16381921
8. Ison J, Rapacki K, Ménager H, Kalaš M, Rydza E, Chmura P, et al. Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res. 2015; 44: D38–47. doi: 10.1093/nar/gkv1116 PMID: 26538599
9. Brazas MD, Yim DS, Yamada JT, Ouellette BFF. The 2011 Bioinformatics Links Directory update: more resources, tools and databases and features to empower the bioinformatics community. Nucleic Acids Res. 2011; 39: W3–7. doi: 10.1093/nar/gkr514 PMID: 21715385
10. Henry VJ, Bandrowski AE, Pepin A-S, Gonzalez BJ, Desfeux A. OMICtools: an informative directory for multi-omic data analysis. Database (Oxford). 2014; 2014: bau069–.
11. Bolser DM, Chibon P-Y, Palopoli N, Gong S, Jacob D, Del Angel VD, et al. MetaBase—the wiki-database of biological databases. Nucleic Acids Res. 2012; 40: D1250–4. doi: 10.1093/nar/gkr1099 PMID: 22139927
12. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2015; 43: D30–5. doi: 10.1093/nar/gku1216 PMID: 25414350
13. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. Methods Mol Biol. 2016; 1374: 23–54. doi: 10.1007/978-1-4939-3167-5_2 PMID: 26519399
14. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 2015; 43: D213–21. doi: 10.1093/nar/gku1243 PMID: 25428371
15. Holliday GL, Bairoch A, Bagos PG, Chatonnet A, Craik DJ, Finn RD, et al. Key challenges for the creation and maintenance of specialist protein resources. Proteins. 2015; 83: 1005–13. doi: 10.1002/prot.24803 PMID: 25820941
16. Babbitt PC, Bagos PG, Bairoch A, Bateman A, Chatonnet A, Chen MJ, et al. Creating a specialist protein resource network: a meeting report for the protein bioinformatics and community resources retreat. Database (Oxford). 2015; 2015: bav063.
17. Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, et al. The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. Nat Biotechnol. 2004; 22: 177–83. doi: 10.1038/nbt926 PMID: 14755292
18. Coordinators NR. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016; 44: D7–D19. doi: 10.1093/nar/gkv1290 PMID: 26615191
19. Gillenwater ZM. Stunning CSS3: A Project-based Guide to the Latest in CSS. 1st ed. New Riders; 2010.
20. Bolchini D, Finkelstein A, Perrone V, Nagl S. Better bioinformatics through usability analysis. Bioinformatics. 2008; 25: 406–412. doi: 10.1093/bioinformatics/btn633 PMID: 19073592

21. Marcus A. Graphic Design for Electronic Documents and User Interfaces. New York, NY, USA: ACM; 1992.
22. Munzner T. Visualization Analysis and Design [Internet]. Florida, USA: CRC Press; 2014. Available: https://www.crcpress.com/Visualization-Analysis-and-Design/Munzner/p/book/9781466508910
23. Pavelin K, Cham JA, de Matos P, Brooksbank C, Cameron G, Steinbeck C. Bioinformatics meets user-centred design: a perspective. PLoS Comput Biol. 2012; 8: e1002554. doi: 10.1371/journal.pcbi.1002554 PMID: 22807660
24. Bastien JMC. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform. 2010; 79: e18–23. doi: 10.1016/j.ijmedinf.2008.12.004 PMID: 19345139
25. Wong B. Color blindness. Nat Methods. 2011; 8: 441. Available: http://www.ncbi.nlm.nih.gov/pubmed/21774112 PMID: 21774112
26. Bostock M, Ogievetsky V, Heer J. $D^3$: Data-Driven Documents. IEEE Trans Vis Comput Graph. 2011; 17: 2301–9. doi: 10.1109/TVCG.2011.185 PMID: 22034350
27. Corpas M, Jimenez R, Carbon SJ, García A, Garcia L, Goldberg T, et al. BioJS: an open source standard for biological visualisation—its status in 2014. F1000Research. 2014; 3: 55. doi: 10.12688/f1000research.3-55.v1 PMID: 25075290
28. Wang R, Perez-Riverol Y, Hermjakob H, Vizcaíno JA. Open source libraries and frameworks for biological data visualisation: A guide for developers. Proteomics. 2014;
29. Smits SA, Ouverney CC. jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. PLoS ONE. 2010; 5: e12267. doi: 10.1371/journal.pone.0012267 PMID: 20805892
30. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics. 2015; btv557.
31. Azam N, Yao J. Comparison of term frequency and document frequency based feature selection metrics in text categorization. Expert Syst Appl. 2012; 39: 4760–4768.
32. Turner B, Razick S, Turinsky AL, Vlasblom J, Crowdy EK, Cho E, et al. iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database (Oxford). 2010; 2010: baq023.
33. Tunkelang D. Faceted Search. Synth Lect Inf Concepts, Retrieval, Serv. Morgan & Claypool Publishers; 2009; 1: 1–80.
34. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. Database (Oxford). 2010; 2010: baq020.
35. O'Connor BD, Day A, Cain S, Arnaiz O, Sperling L, Stein LD. GMODWeb: a web framework for the Generic Model Organism Database. Genome Biol. 2008; 9: R102. doi: 10.1186/gb-2008-9-6-r102 PMID: 18570664
36. Stajich JE, Lapp H. Open source tools and toolkits for bioinformatics: significance, and where are we? Brief Bioinform. 2006; 7: 287–96. doi: 10.1093/bib/bbl026 PMID: 16899494
37. Inc. G. Search Engine Optimization Starter Guide. In: Google Inc. [Internet]. 2010 [cited 29 Jul 2015]. Available: http://static.googleusercontent.com/media/www.google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf
38. Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. Ten simple rules for effective online outreach. PLoS Comput Biol. Public Library of Science; 2015; 11: e1003906. doi: 10.1371/journal.pcbi.1003906 PMID: 25879439
39. Wren JD. URL decay in MEDLINE—a 4-year follow-up study. Bioinformatics. 2008; 24: 1381–5. doi: 10.1093/bioinformatics/btn127 PMID: 18413326
40. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012; 28: 2520–2. doi: 10.1093/bioinformatics/bts480 PMID: 22908215
41. Khare R, Good BM, Leaman R, Su AI, Lu Z. Crowdsourcing in biomedicine: challenges and opportunities. Brief Bioinform. 2015; bbv021–.

EDITORIAL

# Ten Simple Rules for Taking Advantage of Git and GitHub

**Yasset Perez-Riverol[1]*, Laurent Gatto[2], Rui Wang[1], Timo Sachsenberg[3], Julian Uszkoreit[4], Felipe da Veiga Leprevost[5], Christian Fufezan[6], Tobias Ternent[1], Stephen J. Eglen[7], Daniel S. Katz[8], Tom J. Pollard[9], Alexander Konovalov[10], Robert M. Flight[11], Kai Blin[12], Juan Antonio Vizcaíno[1]***

**1** European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, **2** Computational Proteomics Unit, Cambridge Systems Biology Centre, University of Cambridge, Cambridge, United Kingdom, **3** Applied Bioinformatics and Department of Computer Science, University of Tübingen, Tübingen, Germany, **4** Medizinisches Proteom-Center, Ruhr-Universität Bochum, Bochum, Germany, **5** Department of Pathology, University of Michigan, Ann Arbor, Michigan, United States of America, **6** Institute of Plant Biology and Biotechnology, University of Münster, Münster, Germany, **7** Centre for Mathematical Sciences, University of Cambridge, Cambridge, United Kingdom, **8** National Center for Supercomputing Applications and Graduate School of Library and Information Science, University of Illinois, Urbana, Illinois, United States of America, **9** MIT Laboratory for Computational Physiology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **10** Centre for Interdisciplinary Research in Computational Algebra, University of St Andrews, St Andrews, United Kingdom, **11** Department of Molecular Biology and Biochemistry, Markey Cancer Center, Resource Center for Stable Isotope-Resolved Metabolomics, University of Kentucky, Lexington, Kentucky, United States of America, **12** The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark

* yperez@ebi.ac.uk (YPR); juan@ebi.ac.uk (JAV)

CrossMark
click for updates

OPEN ACCESS

**Citation:** Perez-Riverol Y, Gatto L, Wang R, Sachsenberg T, Uszkoreit J, Leprevost FdV, et al. (2016) Ten Simple Rules for Taking Advantage of Git and GitHub. PLoS Comput Biol 12(7): e1004947. doi:10.1371/journal.pcbi.1004947

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** July 14, 2016



**Funding:** This study was supported by Wellcome Trust [grant number WT101477MA] (http://www.wellcome.ac.uk/), BBSRC [grant numbers BB/K01997X/1, BB/I00095X/1, BB/L024225/1 and BB/L002817/1] (http://www.bbsrc.ac.uk/), BMBF grant de.NBI - German Network for Bioinformatics Infrastructure (FKZ031 A 534A) (https://www.denbi.de/), NIH grant numbers R01-GM-094231 and R01-EB-017205 (http://www.nih.gov/), EPSRC [reference EP/M022641/1] (https://www.epsrc.ac.uk), NSF grant number 1252893 (http://www.nsf.gov/), and Novo Nordisk Foundation (http://www.novonordiskfonden.dk/en). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

Bioinformatics is a broad discipline in which one common denominator is the need to produce and/or use software that can be applied to biological data in different contexts. To enable and ensure the replicability and traceability of scientific claims, it is essential that the scientific publication, the corresponding datasets, and the data analysis are made publicly available [1,2]. All software used for the analysis should be either carefully documented (e.g., for commercial software) or, better yet, openly shared and directly accessible to others [3,4]. The rise of openly available software and source code alongside concomitant collaborative development is facilitated by the existence of several code repository services such as SourceForge, Bitbucket, GitLab, and GitHub, among others. These resources are also essential for collaborative software projects because they enable the organization and sharing of programming tasks between different remote contributors. Here, we introduce the main features of GitHub, a popular web-based platform that offers a free and integrated environment for hosting the source code, documentation, and project-related web content for open-source projects. GitHub also offers paid plans for private repositories (see Box 1) for individuals and businesses as well as free plans including private repositories for research and educational use.

GitHub relies, at its core, on the well-known and open-source version control system Git, originally designed by Linus Torvalds for the development of the Linux kernel and now developed and maintained by the Git community. One reason for GitHub's success is that it offers more than a simple source code hosting service [5,6]. It provides developers and researchers with a dynamic and collaborative environment, often referred to as a social coding platform, that supports peer review, commenting, and discussion [7]. A diverse range of efforts, ranging from individual to large bioinformatics projects, laboratory repositories, as well as global

**Competing Interests:** The authors have no affiliation with GitHub, nor with any other commercial entity mentioned in this article. The views described here reflect their own views without input from any third party organization.

### Box 1

By default, GitHub repositories are freely visible to all. Many projects decide to share their work publicly and openly from the start of the project in order to attract visibility and to benefit from contributions from the community early on. Some other groups prefer to work privately on projects until they are ready to share their work. Private repositories ensure that work is hidden but also limit collaborations to just those users who are given access to the repository. These repositories can then be made public at a later stage, such as, for example, upon submission, acceptance, or publication of corresponding journal articles. In some cases, when the collaboration was exclusively meant to be private, some repositories might never be made publicly accessible.

collaborations, have found GitHub to be a productive place to share code and ideas and to collaborate (see Table 1).

Some of the recommendations outlined below are broadly applicable to repository hosting services. However, our main aim is to highlight specific GitHub features. We provide a set of recommendations that we believe will help the reader to take full advantage of GitHub's features for managing and promoting projects in bioinformatics as well as in many other research domains. The recommendations are ordered to reflect a typical development process: learning Git and GitHub basics, collaboration, use of branches and pull requests, labeling and tagging of code snapshots, tracking project bugs and enhancements using issues, and dissemination of the final results.

## Rule 1: Use GitHub to Track Your Projects

The backbone of GitHub is the distributed version control system Git. Every change, from fixing a typo to a complete redesign of the software, is tracked and uniquely identified. Although

**Table 1. Bioinformatics repository examples with good practices of using GitHub.** The table contains the name of the repository, the type of example (issue tracking, branch structure, unit tests), and the URL of the example. All URLs are prefixed with https://github.com/.

| Name of the Repository | Type | URL |
|---|---|---|
| Adam | Community Project, Multiple forks | https://github.com/bigdatagenomics/adam |
| BioPython [18] | Community Project, Multiple contributors | https://github.com/biopython/biopython/graphs/contributors |
| Computational Proteomics Unit | Lab Repository | https://github.com/ComputationalProteomicsUnit |
| Galaxy Project [19] | Community Project, Bioinformatics Repository | https://github.com/galaxyproject/galaxy |
| GitHub Paper | Manuscript, Issue discussion, Community Project | https://github.com/ypriverol/github-paper |
| MSnbase [20] | Individual project repository | https://github.com/lgatto/MSnbase/ |
| OpenMS [21] | Bioinformatics Repository, Issue discussion, branches | https://github.com/OpenMS/OpenMS/issues/1095 |
| PRIDE Inspector Toolsuite [22] | Project Organization, Multiple projects | https://github.com/PRIDE-Toolsuite |
| Retinal wave data repository [23] | Individual project, Manuscript, Binary Data organized | https://github.com/sje30/waverepo |
| SAMtools [24] | Bioinformatics Repository, Project Organization | https://github.com/samtools |
| rOpenSci | Community Project, Issue discussion | https://github.com/ropensci |
| The Global Alliance For Genomics and Health | Community Project | https://github.com/ga4gh |

doi:10.1371/journal.pcbi.1004947.t001

Git has a complex set of commands and can be used for rather complex operations, learning to apply the basics requires only a handful of new concepts and commands and will provide a solid ground to efficiently track code and related content for research projects. Many introductory and detailed tutorials are available (see Table 2 below for a few examples). In particular, we recommend *A Quick Introduction to Version Control with Git and GitHub* by Blischak et al. [5].

In a nutshell, initializing a (local) repository (often abbreviated as *repo*) marks a directory as one to be tracked (Fig 1). All or parts of its content can be added explicitly to the list of files to track.

```
cd project ## move into directory to be tracked
git init ## initialize local repository
## add individual files such as project description, reports,
source code
git add README project.md code.R
git commit -m "initial commit" ## saves the current local snapshot
```

Subsequently, every change to the tracked files, once committed, will be recorded as a new revision, or *snapshot*, uniquely identifying the changes in all the modified files. Git is remarkably effective and efficient in archiving the complete history of a project by, among other things, storing only the differences between files.

In addition to local copies of the repository, it is straightforward to create remote repositories on GitHub (called origin, with default branch master—see below) using the web interface, and then synchronize local and remote repositories.

```
git push origin master ## push local changes to the remote
repository
git pull origin master ## pull remote changes into the local
repository
```

Following Tony Rossini's advice in 2005 to "commit early, commit often, and commit in a repository from which we can easily roll-back your mistakes," one can organize one's work in small incremental changes. At any time, it is possible to go back to a previous version. In larger projects, multiple users are able to work on the same remote repository, with all contributions being recorded, restorable, and attributed to the author.

Users usually track source code, text files, images, and small data files inside their repositories and generally do not track derived files such as build logs or compiled binaries (read Box 2 to see how to handle large binary files in GitHub). And, although the majority of GitHub repositories are used for software development, users can also keep text documents such as analysis

**Table 2. Online courses, tutorials, and workshops about GitHub and Git for scientists.**

| Name of the Material | URL |
|---|---|
| Git help and Git help -a | Document, installed with Git |
| Karl Broman's Git/Github Guide | http://kbroman.org/github_tutorial/ |
| Version Control with GitVersion Control with Git | http://swcarpentry.github.io/git-novice/ |
| Introduction to Git | http://git-scm.com/book/ch1-3.html |
| Github Training | https://training.github.com/ |
| Github Guides | https://guides.github.com/ |
| Good Resources for Learning Git and GitHub | https://help.github.com/articles/good-resources-for-learning-git-and-github/ |
| Software Carpentry: Version Control with Git | http://swcarpentry.github.io/git-novice/ |

doi:10.1371/journal.pcbi.1004947.t002

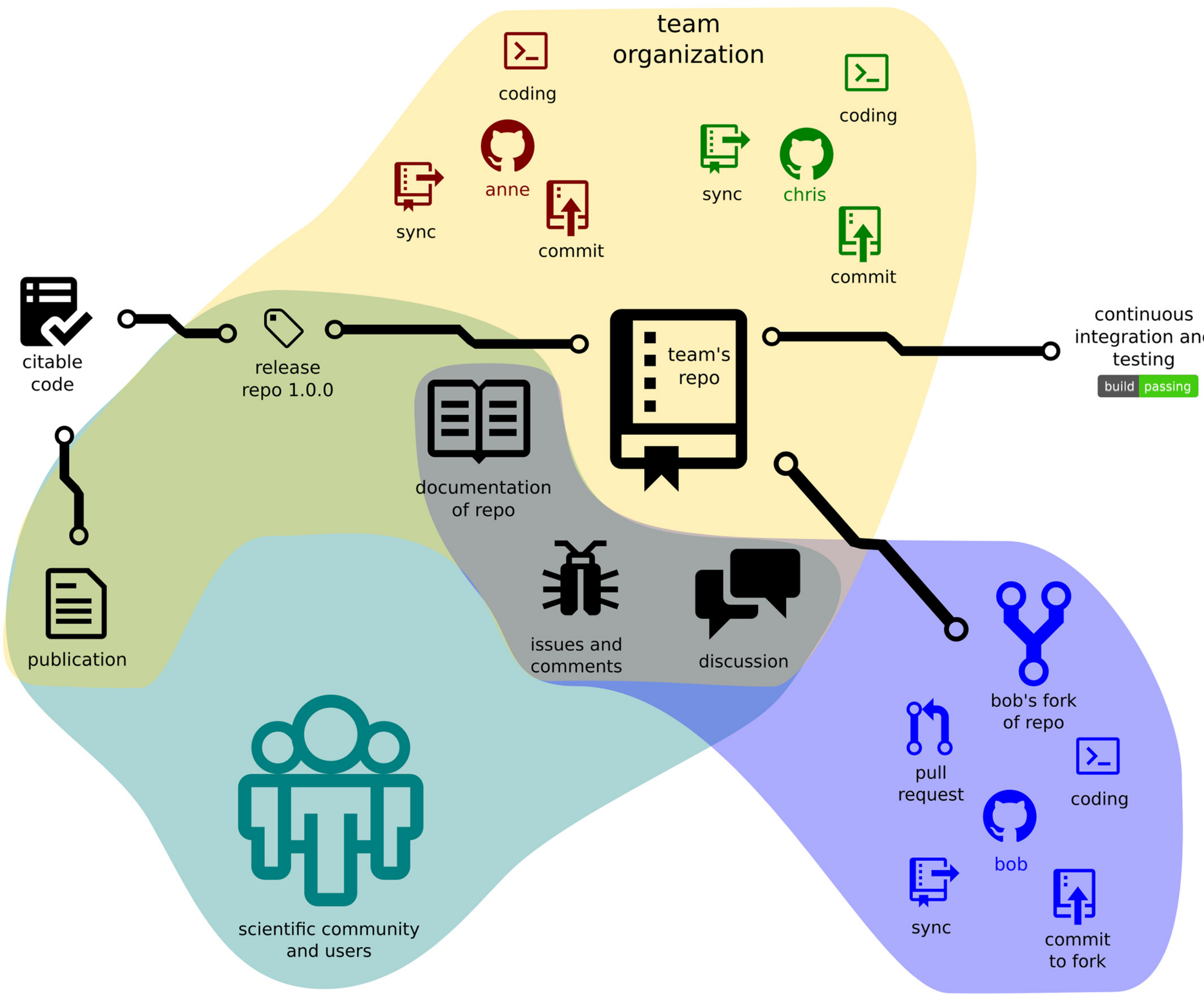



**Fig 1. The structure of a GitHub-based project illustrating project structure and interactions with the community.**

doi:10.1371/journal.pcbi.1004947.g001

reports and manuscripts (see, for example, the repository for this manuscript at https://github.com/ypriverol/github-paper).

Due to its distributed design, each up-to-date local Git repository is an entire exact historical copy of everything that was committed—file changes, commit message logs, etc. These copies act as independent backups as well, present on each user's storage device. Git can be considered to be fault-tolerant because of this, which is a win over centralized version control systems. If the remote GitHub server is unavailable, collaboration and work can continue between users, as opposed to centralized alternatives.

The web interface offered by GitHub provides friendly tools to perform many basic operations and a gentle introduction to a more rich and complex set of functionalities. Various graphical user-interface-driven clients for managing Git and GitHub repositories are also available (https://www.git-scm.com/downloads/guis). Many editors and development environments

### Box 2

Using GitHub or any similar versioning/tracking system is not a replacement for good project management; it is an extension, an improvement for good project and file managing (see for example [9]). One practical consideration when using GitHub, for example, is dealing with large binary files. Binary files such as images, videos, executable files, or many raw data used in bioinformatics, are stored as a single large entity in Git. As a result, every change, even if minimal, leads to a complete new copy of the file in the repository, producing large size increments and the inability to search (see https://help.github.com/articles/searching-code/) and compare file content across revisions. Git offers a Large File Storage (LFS) module that replaces such large files with pointers while the large binary file can be stored remotely, which results in smaller and faster repositories. Git LFS is also supported by GitHub, albeit with a space quota or for a fee, to retain your usual GitHub workflow (https://help.github.com/categories/managing-large-files/) (S1 File, Section 1).

such as, for example, the popular RStudio editor for the R programming language [8], directly integrate with code versioning using Git and GitHub. In addition, for remote Git repositories, GitHub provides its own features that will be described in subsequent rules (Fig 1).

## Rule 2: GitHub for Single Users, Teams, and Organizations

Public projects on GitHub are visible to everyone, but write permission, i.e., the ability to directly modify the content of a repository, needs to be granted explicitly. As a repository owner, you can grant this right to other GitHub users. In addition to being owned by users, repositories can also be created and managed as part of teams and organizations.

Project managers can structure projects to manage permissions at different levels: users, teams, and organizations. Users are the central element of GitHub as in any other social network. Every user has a profile listing their GitHub projects and activities, which can optionally be populated with personal information including name, email address, image, and webpage. To stay up to date with the activity of other users, one can *follow* their accounts (see also Rule 10). Collaboration can be achieved by simply adding a trusted *Collaborator*, thereby granting write access.

However, development in large projects is usually done by teams of people within a larger organization. GitHub organizations are a great way to manage team-based access permissions for the individual projects of institutes, research labs, and large open-source projects that need multiple owners and administrators (Fig 1). We recommend that you, as an individual researcher, make your profile visible to other users and display all of the projects and organizations you are working in.

## Rule 3: Developing and Collaborating on New Features: Branching and Forking

Anyone with a GitHub account can *fork* any repository they have access to. This will create a complete copy of the content of the repository, while retaining a link to the original "upstream" version. One can then start working on the same code base in one's own fork (https://help.github.com/articles/fork-a-repo/) under their username (see, for example, https://github.com/ypriverol/github-paper/network/members for this work) or organization (see Rule 2). Forking a repository allows users to freely experiment with changes without affecting the original project and forms the basis of social coding. It allows anyone to develop and test novel features with existing code and offers the possibility of contributing novel features, bug fixes, and improvements to documentation back into the original upstream project (requested by opening an *pull request*) repository and becoming a contributor. Forking a repository and providing pull requests constitutes a simple method for collaboration inside loosely defined teams and over more formal organizational boundaries, with the original repository owner(s) retaining control over which external contributions are accepted. Once a pull request is opened for review and discussion, it usually results in additional insights and increased code quality [7].

Many contributors can work on the same repository at the same time without running into edit conflicts. There are multiple strategies for this, and the most common way is to use Git *branches* to separate different lines of development. Active development is often performed on a development branch and stable versions, i.e., those used for a software release, are kept in a master or release branch (see for example https://github.com/OpenMS/OpenMS/branches). In practice, developers often work concurrently on one or several features or improvements. To keep commits of the different features logically separated, distinct branches are typically used. Later, when development is complete and verified to work (i.e., none of the tests fail, see Rule 5), new features can be merged back into the development line or master branch. In addition, one can always pull the currently up-to-date master branch into a feature branch to adapt the feature to the changes in the master branch.

When developing different features in parallel, there is a risk of applying incompatible changes in different branches/forks; these are said to become *out of sync*. Branches are just short-term departures from master. If you pull frequently, you will keep your copy of the repository up to date and you will have the opportunity to merge your changed code with others' contributors, ideally without requiring you to manually address conflicts to bring the branches in sync again.

## Rule 4: Naming Branches and Commits: Tags and Semantic Versions

Tags can be used to label versions during the development process. Version numbering should follow "semantic versioning" practice, with the format X.Y.Z., with X being the major, Y the minor, and Z the patch version of the release, including possible meta information, as described in http://semver.org/. This semantic versioning scheme provides users with coherent version numbers that document the extent (bug fixes or new functionality) and backwards compatibility of new releases. Correct labeling allows developers and users to easily recover older versions, compare them, or simply use them to reproduce results described in publications (see Rule 8). This approach also help to define a coherent software publication strategy.

## Rule 5: Let GitHub Do Some Tasks for You: Integrate

The first rule of software development is that the code needs to be ready to use as soon as possible [10], to remain so during development, and that it should be well-documented and tested. In 2005, Martin Fowler defined the basic principles for continuous integration in software development [11]. These principles have become the main reference for best practices in continuous integration, providing the framework needed to deploy software and, in some way, also data. In addition to mere error-free execution, dedicated code testing is aimed at detecting possible bugs introduced by new features or changes in the code or dependencies, as well as detecting wrong results, often known as *logic errors*, in which the source code produces a different result than what was intended. Continuous integration provides a way to automatically and systematically run a series of tests to check integrity and performance of code, a task that can be automated through GitHub.

GitHub offers a set of *hooks* (automatically executed scripts) that are run after each push to a repository, making it easier to follow the basic principles of continuous integration. The GitHub web hooks allow third-party platforms to access and interact with a GitHub repository and thus to automate post-processing tasks. Continuous integration can be achieved by Travis CI, a hosted continued integration platform that is free for all open-source projects. Travis CI builds and tests the source code using a plethora of options such as different platforms and interpreter versions (S1 File, Section 2). In addition, it offers notifications that allow your team and contributors to know if the new changes work and to prevent the introduction of errors in the code (for instance, when merging pull requests), making the repository always ready to use.

## Rule 6: Let GitHub Do More Tasks for You: Automate

More than just code compilation and testing can be integrated into your software project: GitHub hooks can be used to automate numerous tasks to help improve the overall quality of your project. An important complement to successful test completion is to demonstrate that the tests sufficiently cover the existing code base. For this, the integration of Codecov is recommended. This service will report how much of the code base and which lines of code are being executed as part of your code tests. The Bioconductor project, for example, highly recommends that packages implement unit testing (S1 File, Section 2) to support developers in their package development and maintenance (http://bioconductor.org/developers/unitTesting-guidelines/) and systematically tests the coverage of all of its packages (https://codecov.io/github/Bioconductor-mirror/). One might also consider generating the documentation upon code/documentation modification (S1 File, Section 3). This implies that your projects provide comprehensive documentation so others can understand and contribute back to them. For Python or C/C++ code, automatic documentation generation can be done using sphinx and subsequently integrated into GitHub using "Read the Docs." All of these platforms will create reports and badges (sometimes called shields) that can be included on your GitHub project page, helping to demonstrate that the content is of high quality and well-maintained.

## Rule 7: Use GitHub to Openly and Collaboratively Discuss, Address, and Close Issues

GitHub *issues* are a great way to keep track of bugs, tasks, feature requests, and enhancements. While classical issue trackers are primarily intended to be used as bug trackers, in contrast, GitHub issue trackers follow a different philosophy: each tracker has its own section in every repository and can be used to trace bugs, new ideas, and enhancements by using a powerful tagging system. The main objective of *issues* in GitHub is promoting collaboration and providing context by using cross-references.

Raising an issue does not require lengthy forms to be completed. It only requires a title and, preferably, at least a short description. Issues have very clear formatting and provide space for optional comments, which allow anyone with a Github account to provide feedback. For example, if the developer needs more information to be able to reproduce a bug, he or she can simply request it in a comment.

Additional elements of issues are (i) color-coded labels that help to categorize and filter issues, (ii) milestones, and (iii) one assignee responsible for working on the issue. They help developers to filter and prioritize tasks and turn an issue tracker into a planning tool for their project.

It is also possible for repository administrators to create issue and pull request templates (https://help.github.com/articles/helping-people-contribute-to-your-project/) (see Rule 3) to customize and standardize the information to be included when contributors open issues. GitHub issues are thus dynamic, and they pose a low entry barrier for users to report bugs and request features. A well-organized and tagged issue tracker helps new contributors and users to understand a project more deeply. As an example, one issue in the OpenMS repository (https://github.com/OpenMS/OpenMS/issues/1095) allowed the interaction of eight developers and attracted more than one hundred comments. Contributors can add figures, comments, and references to other issues and pull requests in the repository, as well as direct references to code.

As another illustration of issues and their generic and wide application, we (https://github.com/ypriverol/github-paper/issues) and others (https://github.com/ropensci/RNeXML/issues/121) used GitHub issues to discuss and comment on changes in manuscripts and address reviewers' comments.

## Rule 8: Make Your Code Easily Citable, and Cite Source Code!

It is a good research practice to ensure permanent and unambiguous identifiers for citable items like articles, datasets, or biological entities such as proteins, genes, and metabolites (see also Box 3). Digital Object Identifiers (DOIs) have been used for many years as unique and unambiguous identifiers for enabling the citation of scientific publications. More recently, a trend has started to mint DOIs for other types of scientific products such as datasets [12] and training materials (for example [13]). A key motivation for this is to build a framework for giving scientists broader credit for their work [14,15] while simultaneously supporting clearer, more persistent ways to cite and track it. Helping to drive this change are funding agencies such as the National Institutes of Health (NIH) and National Science Foundation (NSF) in the United States and Research Councils in the United Kingdom, which are increasingly recognizing the importance of research products such as publicly available datasets and software.

A common issue with software is that it normally evolves at a different speed than text published in the scientific literature. In fact, it is common to find software having novel features and functionality that were not described in the original publication. GitHub now integrates with archiving services such as Zenodo and Figshare, enabling DOIs to be assigned to code repositories. The procedure is relatively straightforward (see https://guides.github.com/activities/citable-code/), requiring only the provision of metadata and a series of administrative steps. By default, Zenodo creates an archive of a repository each time a new release is created in

### Box 3

Every repository should ideally have the following three files. The first and arguably most important file in a repository is a LICENCE file (see also Rule 8) that clearly defines the permissions and restrictions attached to the code and other files in your repository. The second important file is a README file, which provides, for example, a short description of the project, a quick start guide, information on how to contribute, a TODO list, and links to additional documentation. Such README files are typically written in markdown, a simple markup language that is automatically rendered on GitHub. Finally, a CITATION file to the repository informs your users how to cite and credit your project.

GitHub, ensuring the cited code remains up to date. Once the DOI has been assigned, it can be added to literature information resources such as Europe PubMed Central [16].

As already mentioned in the introduction, reproducibility of scientific claims should be enabled by providing the software, the datasets, and the process leading to interpretable results that were used in a particular study. As much as possible, publications should highlight that the code is freely available in, for example, GitHub, together with any other relevant outputs that may have been deposited. In our experience, this openness substantially increases the chances of getting the paper accepted for publication. Journal editors and reviewers receive the opportunity to reproduce findings during the manuscript review process, increasing confidence in the reported results. In addition, once the paper is published, your work can be reproduced by other members of the scientific community, which can increase citations and foster opportunities for further discussion and collaboration.

The availability of a public repository containing the source code does not make the software open-source per se. You should use an Open Source Initiative (OSI)-approved license that defines how the software can be freely used, modified, and shared. Common licenses such as those listed on http://choosealicense.com are preferred. Note that the LICENSE file in the repository should be a plain-text file containing the contents of an OSI-approved license, not just a reference to the license.

## Rule 9: Promote and Discuss Your Projects: Web Page and More

The traditional way to promote scientific software is by publishing an associated paper in the peer-reviewed scientific literature, though, as pointed out by Buckheir and Donoho, this is just advertising [17]. Additional steps can boost the visibility of an organization. For example, GitHub *Pages* are simple websites freely hosted by GitHub. Users can create and host blog websites, help pages, manuals, tutorials, and websites related to specific projects. Pages comes with a powerful static site generator called Jekyll that can be integrated with other frameworks such as Bootstrap or platforms such as Disqus to support and moderate comments.

In addition, several real-time communication platforms have been integrated with GitHub such as Gitter and Slack. Real-time communication systems allow the user community, developers, and project collaborators to exchange ideas and issues and to report bugs or get support. For example, Gitter is a GitHub-based chat tool that enables developers and users to share aspects of their work. Gitter inherits the network of social groups operating around GitHub repositories, organizations, and issues. It relies on identities within GitHub creating Internet Relay Chat (IRC)-like chat rooms for public and private projects. Within a Gitter chat, members can reference issues, comments, and pull requests. GitHub also supports wikis (which are version-controlled repositories themselves) for each repository, in which users can create and edit pages for documentation, examples, or general support.

A different service is Gist, which represents a unique way to share *code snippets*, single files, parts of files, or full applications. Gists can be generated in two different ways: public gists that can be browsed and searched through Discover and secret gists that are hidden from search engines. One of the main features of Gist is the possibility of embedding code snippets in other applications, enabling users to embed gists in any text field that supports JavaScript.

## Rule 10: Use GitHub to Be Social: Follow and Watch

In the same way researchers are following developments in their field, scientific programmers could follow publicly available projects that might benefit their research. GitHub enables this functionality by *following* other GitHub users (see also Rule 2) or *watching* the activity of projects, which is a common feature in many social media platforms. Take advantage of it as much as possible!

## Conclusions

If you are involved in scientific research and have not used Git and GitHub before, we recommend that you explore its potential as soon as possible. As with many tools, a learning curve lays ahead, but several basic yet powerful features are accessible even to the beginner and may be applied to many different use-cases [6]. We anticipate the reward will be worth your effort. To conclude, we would like to recommend some examples of bioinformatics repositories in GitHub (Table 1) and some useful training materials, including workshops, online courses, and manuscripts (Table 2).

## Supporting Information

**S1 File. Supplementary Information including three sections: Git Large File Storage (LFS), Testing Levels of the Source Code and Continuous integration, and Source code documentation.**
(PDF)

## Acknowledgments

The authors would like to thank C. Titus Brown for mentioning the manuscript on social media, leading to additional contributions and further improvements. We also thank Peter Cock (peterjc) for helpful suggestions contributed directly through GitHub.

## References

1. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten simple rules for the care and feeding of scientific data. PLoS Comput Biol. 2014; 10(4):e1003542. doi: 10.1371/journal.pcbi.1003542 PMID: 24763340
2. Perez-Riverol Y, Alpi E, Wang R, Hermjakob H, Vizcaíno JA. Making proteomics data accessible and reusable: current state of proteomics databases and repositories. Proteomics. 2015; 15(5–6):930–49. doi: 10.1002/pmic.201400302 PMID: 25158685
3. Osborne JM, Bernabeu MO, Bruna M, Calderhead B, Cooper J, Dalchau N, et al. Ten simple rules for effective computational research. PLoS Comput Biol. 2014; 10(3):e1003506. doi: 10.1371/journal.pcbi.1003506 PMID: 24675742
4. Vihinen M. No more hidden solutions in bioinformatics. Nature. 2015; 521(7552):261. doi: 10.1038/521261a PMID: 25993922
5. Blischak J, Davenport E, Wilson G. A Quick Introduction to Version Control with Git and GitHub. PLoS Comput Biol. 2016; 12(1):e1004668. doi: 10.1371/journal.pcbi.1004668 PMID: 26785377
6. Ram K. Git can facilitate greater reproducibility and increased transparency in science. Source code for biology and medicine. 2013; 8(1):1.
7. Dabbish L, Stuart C, Tsay J, Herbsleb J. Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. CSCW '12. New York, NY, USA: ACM; 2012. p. 1277–1286. Available from: http://doi.acm.org/10.1145/2145204.2145396.
8. R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: https://www.R-project.org/.
9. Noble WS. A quick guide to organizing computational biology projects. PLoS Comput Biol. 2009; 5(7): e1000424. doi: 10.1371/journal.pcbi.1000424 PMID: 19649301
10. Leprevost FV, Barbosa VC, Francisco EL, Perez-Riverol Y, Carvalho PC. On best practices in the development of bioinformatics software. Front Genet. 2014; 5:199. doi: 10.3389/fgene.2014.00199 PMID: 25071829
11. Fowler M. Continuous Integration; 2006. http://www.martinfowler.com/articles/continuousIntegration.html.

12. Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol. 2014; 32(3):223–6. doi: 10.1038/nbt.2839 PMID: 24727771

13. Ahmadia A, Aiello-Lammens M, Ainsley J, Allen J, Alsheikh-Hussain A, Banaszkiewicz P, et al. Software Carpentry: Programming with R; 2015. http://dx.doi.org/10.5281/zenodo.27353.

14. Credit where credit is overdue. Nat Biotechnol. 2009; 27(7):579. doi: 10.1038/nbt0709-579 PMID: 19587644

15. FORCE11 Software Citation Working Group. https://www.force11.org/group/software-citation-working-group.

16. Europe PMC Consortium. Europe PMC: a full-text literature database for the life sciences and platform for innovation. Nucleic Acids Res. 2015; 43(Database issue):D1042–8. doi: 10.1093/nar/gku1061 PMID: 25378340

17. Buckheit J, Donoho D. WaveLab and Reproducible Research. Springer-Verlag; 1995. p. 55–81.

18. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25(11):1422–3. doi: 10.1093/bioinformatics/btp163 PMID: 19304878

19. Goecks J, Nekrutenko A, Taylor J, Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010; 11(8):R86. doi: 10.1186/gb-2010-11-8-r86 PMID: 20738864

20. Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics. 2012; 28(2):288–289. doi: 10.1093/bioinformatics/btr645 PMID: 22113085

21. Sturm M, Bertsch A, Gröpl C, Hildebrandt A, Hussong R, Lange E, et al. OpenMS—an open-source software framework for mass spectrometry. BMC bioinformatics. 2008; 9(1):163.

22. Perez-Riverol Y, Xu QW, Wang R, Uszkoreit J, Griss J, Sanchez A, et al. PRIDE Inspector Toolsuite: Moving Toward a Universal Visualization Tool for Proteomics Data Standard Formats and Quality Assessment of ProteomeXchange Datasets. Molecular & Cellular Proteomics. 2016; 15(1):305–317. doi: 10.1074/mcp.O115.050229

23. Eglen SJ, Weeks M, Jessop M, Simonotto J, Jackson T, Sernagor E. A data repository and analysis framework for spontaneous neural activity recordings in developing retina. Gigascience. 2014; 3(1):3. doi: 10.1186/2047-217X-3-3 PMID: 24666584

24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. doi: 10.1093/bioinformatics/btp352 PMID: 19505943

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for the Open Development of Scientific Software

**Andreas Prlić[1]*, James B. Procter[2]**

**1** San Diego Supercomputer Center, University of California San Diego, La Jolla, California, United States of America, **2** School of Life Sciences Research, College of Life Sciences, University of Dundee, Dundee, Scotland, United Kingdom

Open-source software development has had significant impact, not only on society, but also on scientific research. Papers describing software published as open source are amongst the most widely cited publications (e.g., BLAST [1,2] and Clustal-W [3]), suggesting many scientific studies may not have been possible without some kind of open software to collect observations, analyze data, or present results. It is surprising, therefore, that so few papers are accompanied by open software, given the benefits that this may bring.

Publication of the source code you write not only can increase your impact [4], but also is essential if others are to be able to reproduce your results. Reproducibility is a tenet of computational science [5], and critical for pipelines employed in data-driven biological research. Publishing the source for the software you created as well as input data and results allows others to better understand your methodology, and why it produces, or fails to produce, expected results. Public release might not always be possible, perhaps due to intellectual property policies at your or your collaborators' institutes; and it is important to make sure you know the regulations that apply to you. Open licensing models can be incredibly flexible and do not always prevent commercial software release [5].

Simply releasing the source under an open license, however, is not sufficient if you wish your code to remain useful beyond its publication [6]. The sustainability of software after publication is probably the biggest problem faced by researchers who develop it, and it is here that participating in open development from the outset can make the biggest impact. Grant-based funding is often exhausted shortly after new software is released, and without support, in-house maintenance of the software and the systems it depends on becomes a struggle. As a consequence, the software will cease to work or become unavailable for download fairly quickly [7], which may contravene archival policies stipulated by your journal or funding body. A collaborative and open project allows you to spread the resource and maintenance load to minimize these risks, and significantly contributes to the sustainability of your software.

If you have the choice, embracing an open approach to development has tremendous benefits. It allows you to build on the work of other scientists, and enables others to build on your own efforts. To make the development of open scientific software more rewarding and the experience of using software more positive, the following ten rules are intended to serve as a guide for any computational scientist.

## Rule 1: Don't Reinvent the Wheel

As in any other field, you should do some research before starting a new programming project to find out if aspects of your problem have already been solved. Many fundamental scientific algorithms and methods have already been implemented in open-source libraries, and having the source means you can easily evaluate if they will work in your situation. You can also contact online communities (see [8]) to find out about their experiences with existing approaches, and if none are appropriate, any new implementation you provide will be well received, however modest. Providing another solution to a problem, even if technologically novel, is only an accomplishment in engineering and rarely suitable for publication on its own. However, if it is useful it can benefit everyone, even if it addresses a mundane task. Furthermore, when there are no existing implementations for your platform, or they cannot cope with the size, complexity, or other specifics of your data, then new approaches may be required that lead to new science.

## Rule 2: Code Well

If you don't know them already, learn the basics of software development [9,10]. You don't need to be the best software developer in the world, but try to be inspired by them. Study other people's code and learn by practice. Join an existing open-source project. There are plenty to choose from (most open-source repositories have a "biology" or "bioinformatics" project tag), but the "bio-*" projects hosted at the Open Bioinformatics Foundation are a good place to start [11–14]. Once you identify a weakness (and you will!) or something that does not work as expected, fix the issue so it works for yourself and provide a patch back to the original authors. Getting familiar with other people's code in this way is a great way to boost your experience and learn new techniques.

## Rule 3: Be Your Own User

One of the more graphic mottos in the open-source community is "eat your own dog food". For a researcher this has two implications. If you are developing software of value to your field, it is important that you demonstrate that it can address important questions in a useful or novel way. The second implication is that your software should be useful to other developers, and is not simply a demonstration of the solution. Sadly, for some scientific

**Citation:** Prlić A, Procter JB (2012) Ten Simple Rules for the Open Development of Scientific Software. PLoS Comput Biol 8(12): e1002802. doi:10.1371/journal.pcbi.1002802

**Published** December 6, 2012



**Funding:** The authors received no specific funding for writing this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: andreas.prlic@gmail.com

Andreas Prlić is a Software Editor for *PLOS Computational Biology*.

software articles this is often not the case, and there are examples of software that—whilst novel—were not developed to solve a problem the scientists faced in a practical situation. Problems to do with how software is structured or functions in a variety of situations are difficult to detect during peer review. It is only later, when a researcher discovers and applies the software during their research, that these issues hinder or obstruct progress. Avoiding wasted effort of this kind is critical to researchers, who have limited time and require high levels of quality and reproducibility from scientific source code. By being "your own best user" many such problems will be detected before they become public.

## Rule 4: Be Transparent

Scientific software, like other competitive activities, is often at first developed behind closed doors instead of out in the open, and public release is then only considered around the time of publication. The first reason given for this (after any legal constraints), is the fear of getting scooped—that somebody else might use the ideas to produce competing software faster or tackle the same research problem first. In our experience, however, open development often results in just the opposite. Founding or contributing new code to open-source projects is one way for a researcher to stake a claim in a field [15]. People with similar or related research interests who discover the project will find that they have more to gain from collaborating than from competing with the original developers. The second reason given for closed development is the perhaps more serious risk that code released prematurely may lead to incorrect findings by others. However, examples regularly show [16] that even prior publication of software in a peer-reviewed journal does not preclude the presence of serious bugs. One consequence of transparent, open development is that it allows many eyes to evaluate the code and recognize and fix any issues, which reduces the likelihood of serious errors in the final product. There are public repositories such as Sourceforge or GitHub that greatly facilitate this kind of team development approach. They provide free services such as version control, Wikis, mailing lists, and bug trackers and support communication with your collaborators to share effort, document bugs, and solve problems more quickly [17]. Several models for initiating and managing open development have also been proposed and advocated by different communities, such as the Apache Way [18,19].

## Rule 5: Be Simple

Science is hard enough already. If your software is too complex to obtain and operate or can only run on one platform, then few people will bother to try it out, and even fewer will use it successfully (particularly your reviewers!). This is doubly important for open projects, since difficult compilation or installation processes will raise a barrier against participation. Documentation helps a lot, in the form of build and installation instructions, user manuals, or even video demonstrations, but simplicity is key, since potential users will first evaluate how long it will take to install and get something out of your software against the time it will take them to find another way. Employ standard package or software installation models for as many platforms as possible. Practically all operating systems, and many languages (e.g., Perl, Ruby, and Python), have standard models for creating installable software packages, which allow you to specify any other software your code needs to run, and make it easier for you to distribute it [20]. If you don't have the time to learn how to create an installation package yourself, then get in contact with one of the many open-source packaging communities (e.g., DebianMed), and ask for help. When creating new software, try to support standard file formats and don't come up with new, custom formats. This can make your software less appealing. Spending time to create online documentation, sample data files, and test cases will give others an easy start into your codebase.

## Rule 6: Don't Be a Perfectionist

Don't wait too long with getting the first version of your source code out into the public and don't worry too much if your first prototypes still have critical features missing. If your idea is innovative, others will understand the concept. Moreover, as scientists, we are trained to constantly assess and revise our own and each others' hypotheses, and we should do the same for our software. "Release early, release often" is regarded as an open-source mantra, and attributed to Linus Torvalds by Eric Raymond [21]. It advocates the practice of releasing as soon as new work has been done, because your "customers" will quickly identify problems and new requirements, and you will be able to fix them more quickly if you avoid sitting on and polishing new code for several months before letting it into the wild. Agile development practices [22], which have become popular in the last decade, embody this iterative development process.

## Rule 7: Nurture and Grow Your Community

The biggest advantage of open development is that it allows users and developers to freely interact and form communities, and if your software is useful, your user base will grow. You can only do so much by yourself, but if you form a team (see [23]) and communicate with the people who use your tool, then new scientific and technical collaborations can arise. Reciprocity is essential, however: as a user of open source, acknowledge the tools you are using. If you are running your own open community, acknowledge the contributions of each person to your project. Make it easy for others to contribute ideas and act on feedback. Seeing that suggestions are being taken seriously and acted upon can be highly motivating and will encourage further involvement. Try to avoid changing key aspects of your code that other people's software or analysis pipelines might depend on, such as file formats, command line arguments, or application programming interfaces (APIs). If you do, discuss them online first, then document and create demonstrations of the changes, and assign a version number to the API. Even better, use Semantic Versioning (http://semver.org), which communicates both API and software version compatibility between releases. Above all, avoid confusing your users—drastic differences between each release that introduce incompatibilities will win no friends.

## Rule 8: Promote Your Project

In order to attract more attention to your project, it is important to spend time promoting it. Appearance matters, and a clean, well-organized website that will help your cause is not hard to achieve. Hosting sites such as GitHub or Google code provide standard templates for project websites, where you only need to come up with a name and logo. Branding is not rocket science, but it is about habit—once you have a name, stick with it, and use it everywhere. Create personae for your project on social networks that people can connect to, and increase your presence in online discussion forums: answer questions on ResearchGate, Linkedin, or any of the other open communities where potential users of your software might be. Whilst doing this, bear in mind that regardless of how good your project is, people are more likely to connect with your project because of what you say and your own personal profile. Finally, remember about more traditional ways of communicating your work: go to conferences where you will meet other developers and potential users of your software, and give as many presentations as you can. Keep an eye out for ad hoc developer meetups and hackathons, where open-source coders get together to work on one, or many different projects. Promotion is hard work, but through it you will grow and strengthen your community.

## Rule 9: Find Sponsors

No matter how large the community around your project and how efficiently it is developed and managed, some level of funding is essential. Scientific software can be successfully supported through grants, by writing applications to address new scientific problems through the development and use of software, or attaching development and upkeep of software as a deliverable on experimental grants. Grant writing [24] is beyond the scope of the Ten Simple Rules presented here, but it is worth mentioning that if the rules laid out here are being followed, an open development community can ensure value beyond the lifetime of an award. Open development directly addresses the section on sustainability in grant applications, but the emphasis here has to be on the community. Simply releasing code openly, without support and maintenance, will not ensure extended value; instead, you need to explain how you will actively foster your community of users and developers. Besides grants, there are also other support models for open source. Internship programs like the Google Summer of Code finance students to spend a summer working on open-source projects, and a number of projects related to science have benefited from them.

## Rule 10: Science Counts

As scientists, the software we write is primarily a means to advance our research and, ultimately, achieve our scientific goals. Whilst the development of software for the consumption of others aligns well with other processes of scientific advancement, it is the science that ultimately counts. Scientific software development fulfils an immediate need, but maintenance of code that is no longer relevant to your own research is a serious time sink, and will rarely lead to your next paper, or secure your next grant or position. Open-source development and maintenance is an intensely social process, and perhaps particularly appealing to scientists since we tend to crave interaction with others as knowledgeable about our fields as ourselves. These aspects of open source make it even more important for us as scientists to keep an eye on the big picture, and stay true to our scientific goals. However, if done right, you can publish both the science and the software for the same project, giving credit to everyone involved. Open-source communities ensure persistence of projects by allowing project leadership to be shared and passed to other members. As a scientist, this offers you the opportunity to naturally progress to new challenges with the knowledge that the software you created will remain available and benefit others.

## Acknowledgments

We want to thank Hilmar Lapp for his comments and contributions to parts of the text. Thanks to Scott Markel, Fran Lewitter, and Spencer Bliven for their comments and feedback.

## References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2
2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402. doi:10.1093/nar/25.17.3389
3. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680. doi:10.1093/nar/22.22.4673
4. Vandewalle P (2012) Code sharing is associated with research impact in image processing. IEEE Computing in Science & Engineering 14: 42–47. http://rr.epfl.ch/37/.
5. Morin A, Urban J, Sliz P (2012) A quick guide to software licensing for the scientist-programmer. PLoS Comput Biol 8(7): e1002598. doi:10.1371/journal.pcbi.1002598
6. Cock P (2011) Opening up NCBI BLAST? Available: http://blastedbio.blogspot.co.uk/2011/08/opening-up-ncbi-blast.html Accessed 27 October 2012.
7. Wren JD (2008) URL decay in MEDLINE—a 4-year follow-up study. Bioinformatics 24(11): 1381–1385. doi:10.1093/bioinformatics/btn127
8. Dall'Olio GM, Marino J, Schubert M, Keys KL, Stefan MI, et al. (2011) Ten simple rules for getting help from online scientific communities. PLoS Comput Biol7(9): e1002202. doi:10.1371/journal.pcbi.1002202
9. Sofware Carpentry (2012) http://software-carpentry.org/ Accessed 27 October 2012.
10. Aruliah DA, Titus Brown C, Chue Hong NP, Davis M, Guy RT, et al. (2012) Best practices for scientific computing. eprint arXiv:1210.0530. Available: http://arxiv.org/abs/1210.0530. Accessed 27 October 2012.
11. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, et al. (2012) The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12(10): 1611–1618. doi:10.1101/gr.361602
12. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11): 1422–1423. Epub 20 Mar 2009. doi:10.1093/bioinformatics/btp163
13. Prlić A, Yates A, Bliven SE, Rose PW, Jacobsen J, et al. (2012) BioJava: an open-source framework for bioinformatics in 2012. Bioinformatics 28(20): 2693–2695. Epub 9 Aug 2012. doi:10.1093/bioinformatics/bts494
14. Goto N, Prins P, Nakao M, Bonnal R, Aerts J, et al. (2010) BioRuby: bioinformatics software for the Ruby programming language. Bioinformatics 26(20): 2617–2619. doi:10.1093/bioinformatics/btq475
15. Barnes N (2012) The science code manifesto. Available: http://sciencecodemanifesto.org/. Accessed 15 October 2012.
16. Zeeya Merali (2010) Computational science: …error…why scientific programming does not compute. Nature 467: 775–777. doi:10.1038/467775a
17. Rother K, Potrzebowski W, Puton T, Rother M, Wywial E, et al. (2012) A toolbox for developing bioinformatics software. Brief Bioinform 13(2): 244–257. Epub 29 Jul 2011. doi:10.1093/bib/bbr035
18. The Apache Software Foundation. The Apache way. Available: http://incubator.apache.org/learn/theapacheway.html. Accessed 17 October 2012.
19. Gardler R (2010) Community development at the Apache Software Foundation. Ross Gardler's keynote from BOSC 2010. Available: http://www.slideshare.net/bosc2010/gardler-bosc2010-communitydevelopmentattheasf. Accessed 27 October 2012.
20. Bonnal RJ, Aerts J, Githinji G, Goto N, MacLean D, et al. (2012) Biogem: an effective tool-based approach for scaling up open source software development in bioinformatics. Bioinformatics 28(7): 1035–1037. Epub 12 Feb 2012. doi:10.1093/bioinformatics/bts080
21. Raymond ES (1999) The cathedral & the bazaar. O'Reilly. ISBN 1-56592-724-9.
22. Kane DW, Hohman MM, Cerami EG, McCormick MW, Kuhlmman KF, et al. (2006) Agile methods in biomedical software development: a multi-site experience report. BMC Bioinformatics 7: 273. doi:10.1186/1471-2105-7-273
23. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3(3): e44. doi:10.1371/journal.pcbi.0030044
24. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comput Biol 2(2): e12. doi:10.1371/journal.pcbi.0020012

EDITORIAL

# Ten simple rules for innovative dissemination of research

**Tony Ross-Hellauer[1]*, Jonathan P. Tennant[2], Viltė Banelytė[3], Edit Gorogh[4], Daniela Luzi[5], Peter Kraker[6], Lucio Pisacane[5], Roberta Ruggieri[5], Electra Sifacaki[7], Michela Vignoli[8]**

**1** Open and Reproducible Research Group, Institute of Interactive Systems and Data Science, Graz University of Technology and Know-Center GmbH, Graz, Austria, **2** Center for Research and Interdisciplinarity, University of Paris, Paris, France, **3** Freelance Researcher, Vilnius, Lithuania, **4** University and National Library, University of Debrecen, Debrecen, Hungary, **5** Institute for Research on Population and Social Policies, National Research Council, Rome, Italy, **6** Open Knowledge Maps, Vienna, Austria, **7** National and Kapodistrian University of Athens, Athens, Greece, **8** Center for Digital Safety and Security, AIT Austrian Institute of Technology, Vienna, Austria

* tross@know-center.at

Check for updates

OPEN ACCESS

**Citation:** Ross-Hellauer T, Tennant JP, Banelytė V, Gorogh E, Luzi D, Kraker P, et al. (2020) Ten simple rules for innovative dissemination of research. PLoS Comput Biol 16(4): e1007704. https://doi.org/10.1371/journal.pcbi.1007704

**Editor:** Russell Schwartz, Carnegie Mellon University, UNITED STATES

**Published:** April 16, 2020



**Funding:** This work was partly funded by the OpenUP project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 710722. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** We have read the journal's policy and have the following conflicts: TR-H is Editor-in-Chief of the open access journal *Publications*. JT is the founder of the Open Science MOOC, and a former employee of ScienceOpen.

## Author summary

How we communicate research is changing because of new (especially digital) possibilities. This article sets out 10 easy steps researchers can take to disseminate their work in novel and engaging ways, and hence increase the impact of their research on science and society.

## Introduction

As with virtually all areas of life, research dissemination has been disrupted by the internet and digitally networked technologies. The last two decades have seen the majority of scholarly journals move online, and scholarly books are increasingly found online as well as in print. However, these traditional communication vehicles have largely retained similar functions and formats during this transition. But digital dissemination can happen in a variety of ways beyond the traditional modes: social media have become more widely used among researchers [1,2,3], and the use of blogs and wikis as a specific form of 'open notebook science' has been popular for more than a decade [4].

Professional academic social networks such as ResearchGate and Academia.edu boast millions of users. New online formats for interaction with the wider public, such as TED talks broadcast via YouTube, often receive millions of views. Some researchers have even decided to make all of their research findings public in real time by keeping open notebooks [5,6]. In particular, digital technologies invoke new ways of reaching and involving audiences beyond their usual primary dissemination targets (i.e., other scholars) to actively involve peers or citizens who would otherwise remain out of reach for traditional methods of communication [7]. Adoption of these outlets and methods can also lead to new cross-disciplinary collaborations, helping to create new research, publication, and funding opportunities [8].

Beyond the increase in the use of web-based and computational technologies, other trends in research cultures have had a profound effect on dissemination. The push towards greater public understanding of science and research since the 1980s, and an emphasis on engagement and participation of non-research audiences have brought about new forms of dissemination [9]. These approaches include popular science magazines and science shows on television and the radio. In recent years, new types of events have emerged that aim at involving the general public within the research process itself, including science slams and open lab days. With science cafés and hackerspaces, novel, participatory spaces for research production and dissemination are emerging—both online and offline. Powerful trends towards responsible research and innovation, the increasing globalisation of research, and the emergence and inclusion of new or previously excluded stakeholders or communities are also reshaping the purposes of dissemination as well as the scope and nature of its audiences.

Many now view wider dissemination and public engagement with science to be a fundamental element of open science [10]. However, there is a paradox at play here, for while there have never been more avenues for the widespread dissemination of research, researchers tend nonetheless to value and focus upon just a few traditional outputs: journal articles, books, and conference presentations [11].

Following Wilson and colleagues [12], we here define research dissemination as a planned process that involves consideration of target audiences, consideration of the settings in which research findings are to be received, and communicating and interacting with wider audiences in ways that will facilitate research uptake and understanding. Innovative dissemination, then, means dissemination that goes beyond traditional academic publishing (e.g., academic journals, books, or monographs) and meetings (conferences and workshops) to achieve more widespread research uptake and understanding. Hence, a citizen science project, which involves citizens in data collection but does not otherwise educate them about the research, is not here considered innovative dissemination.

We here present 10 steps researchers can take to embrace innovative dissemination practices in their research, either as individuals or groups (Fig 1). They represent the synthesis of multidimensional research activities undertaken within the OpenUP project (https://www.openuphub.eu/). This European Coordination and Support Action grant award addressed key aspects and challenges of the currently transforming science landscape and proposed recommendations and solutions addressing the needs of researchers, innovators, the public, and funding bodies. The goal is to provide stakeholders (primarily researchers but also intermediaries) with an entry point to innovative dissemination, so that they can choose methods and tools based on their audience, their skills, and their requirements. The advice is directed towards both individual researchers and research teams or projects. It is similar to other entries in the Ten Simple Rules series (e.g., [13,14]). Ultimately, the benefit here for researchers is increased recognition and social impact of their work.

## Rule 1: Get the basics right

Despite changes in communication technologies and models, there are some basic organisational aspects of dissemination that remain important: to define objectives, map potential target audience(s), target messages, define mode of communication/engagement, and create a dissemination plan. These might seem a bit obvious or laborious but are critical first steps towards strategically planning a project.

### Define objectives

The motivation to disseminate research can come in many forms. You might want to share your findings with wider nonacademic audiences to raise awareness of particular issues or invite audience engagement, participation, and feedback. Start by asking yourself what you want to achieve with your dissemination. This first strategic step will make all other subsequent steps much simpler, as well as guide how you define the success of your activities.


**Fig 1. Summary of the 10 simple rules presented in this paper.**

https://doi.org/10.1371/journal.pcbi.1007704.g001

## Map your audience

Specify who exactly you want your research results to reach, for which purposes, and what their general characteristics might be (e.g., policy makers, patient groups, non-governmental organisations). Individuals are not just 'empty vessels' to be filled with new knowledge, and having a deeper contextual understanding of your audience can make a real difference to the success of your engagement practices. Who is most affected by your research? Who might find it most valuable? What is it that you want them to take away? Get to know your target audiences, their needs and expectations of the research outcomes, as well as their preferred communication channels to develop a detailed understanding of their interests and align your messages and media with their needs and priorities. Keep in mind, too, that intermediaries such as journalists or science communication organisations can support or mediate the dissemination process.

### Target/frame your messages

Target and frame the key messages that you want to communicate to specific groups. Think first from the perspective of what they might want or need to hear from you, rather than what you want to tell them. Choosing media and format of your communication strongly depends on your communication objectives, i.e., what you want to achieve. There are many ways to communicate your research; for example, direct messages, blog/vlog posts, tweeting about it, or putting your research on Instagram. Form and content go hand in hand. Engage intermediaries and leverage any relevant existing networks to help amplify messages.

### Create a dissemination plan

Many funded research projects require a dissemination plan. However, even if not, the formal exercise of creating a plan at the outset that organises dissemination around distinct milestones in the research life cycle will help you to assign roles, structure activities, as well as plan funds to be allocated in your dissemination. This will ultimately save you time and make future work easier. If working in groups, distribute tasks and effort to ensure regular updates of content targeted to different communities. Engage those with special specific skills in the use and/or development of appropriate communication tools, to help you in using the right language and support you in finding the suitable occasions to reach your identified audience. Research is not linear, however, and so you might find it best to treat the plan as a living document to be flexibly adapted as the direction of research changes.

## Rule 2: Keep the right profile

Whether communicating as an individual researcher, a research project, or a research organisation, establishing a prominent and unique identity online and offline is essential for communicating. Use personal websites, social media accounts, researcher identifiers, and academic social networks to help make you and your research visible. When doing this, try to avoid any explicit self-promotion—your personal profile naturally will develop based on your ability to be an effective and impactful communicator.

Academia is a prestige economy, where individual researchers are often evaluated based on their perceived esteem or standing within their communities [15]. Remaining visible is an essential part of accumulating esteem. An online presence maintained via personal websites, social media accounts (e.g., Facebook, Twitter, LinkedIn), researcher identifiers (e.g., ORCID), and academic social networks (e.g., ResearchGate, institutional researcher profiles) can be a personal calling card, where you can highlight experience and demonstrate your expertise in certain topics. Being active on important mailing lists, forums, and social media is not only a good chance to disseminate your findings to those communities but also offers you the chance to engage with your community and potentially spark new ideas and collaborations.

Using researcher identifiers like ORCID when disseminating outputs will ensure that those outputs will be unambiguously linked back to the individual researcher (and even automatically updated to their ORCID profile). The OpenUP survey showed that nearly half of the respondents (41%) use academic social networks as a medium to disseminate their research, and a quarter of respondents (26%) said that these networks informed their professional work [16].

Create a brand by giving your project a unique name, ideally with some intuitive relation to the issue you are investigating. Create a striking visual identity, with a compelling logo, core colours, and a project slogan. Create a website that leverages this visual identity and is as simple and intuitive as possible, both in its layout and in the way content is formulated (limit insider jargon). Create associated appropriate social media accounts (e.g., Twitter, Facebook, LinkedIn, SlideShare, YouTube) and link to this from the project website. Aim for a sustained presence with new and engaging content to reinforce project messaging, and this can help to establish a core following group or user base within different platforms. Include links to other project online presences such as social media accounts, or a rolling feed of updates if possible. Consider including a blog to disseminate core findings or give important project updates. A periodical newsletter could be released in order to provide project updates and other news, to keep the community informed and activated regarding project issues. Depending on the size of your project and budget, you might want to produce hard copy material such as leaflets or fact sheets, as well as branded giveaways to increase awareness of your project. Finally, and perhaps most importantly, try not to come across as a 'scientific robot', and make sure to communicate the more human personality side of research.

## Rule 3: Encourage participation

In the age of open research, don't just broadcast. Invite and engage others to foster participation and collaboration with research audiences. Scholarship is a collective endeavour, and so we should not expect its dissemination to be unidirectional, especially not in the digital age. Dissemination is increasingly done at earlier stages of the research life cycle, and such wider and more interactive engagement is becoming an integral part of the whole research workflow.

Such participative activities can be as creative as you wish; for example, through games, such as Foldit for protein folding (https://fold.it/portal/). You might even find it useful to actively engage 'citizen scientists' in research projects; for example, to collect data or analyse findings. Initiatives such as Zooniverse (https://www.zooniverse.org/) serve as great examples of allowing anyone to freely participate in cutting-edge 'people-powered research'.

Disseminating early and often showcases the progress of your work and demonstrates productivity and engagement as part of an agile development workflow. People like to see progress and react positively to narrative, so give regular updates to followers on social media, for example, blogging or tweeting early research findings for early feedback. Alternatively, involving businesses early on can align research to industry requirements and expectations, thus potentially increasing commercial impact. In any case, active involvement of citizens and other target audiences beyond academia can help increase the societal impact of your research [17].

## Rule 4: Open science for impact

Open science is 'transparent and accessible knowledge that is shared and developed through collaborative networks', as defined by one systematic review [18]. It encompasses a variety of practices covering a range of research processes and outputs, including areas like open access (OA) to publications, open research data, open source software/tools, open workflows, citizen science, open educational resources, and alternative methods for research evaluation including open peer review [19]. Open science is rooted in principles of equitable participation and transparency, enabling others to collaborate in, contribute to, scrutinise and reuse research, and spread knowledge as widely as possible [20]. As such, innovative dissemination is a core element of open science.

Embracing open science principles can boost the impact of research. Firstly, OA publications seem to accrue more citations than their closed counterparts, as well as having a variety of possible wider economic and societal benefits [21]. There are a number of ways to make research papers OA, including at the journal site itself, or self-archiving an accepted manuscript in a repository or personal website.

Disseminating publications as preprints in advance of or parallel to journal submission can increase impact, as measured by relative citation counts [22]. Very often, traditional publishing takes a long time, with the waiting time between submission and acceptance of a paper being in excess of 100 days [23]. Preprinting speeds up dissemination, meaning that findings are available sooner for sharing and reuse. Potential platforms for disseminating preprints include the Open Science Framework, biorXiv, or arXiv.

Dissemination of other open science outputs that would usually remain hidden also not only helps to ensure the transparency and increased reproducibility of research [24], but also means that more research elements are released that can potentially impact upon others by creating network effects through reuse. Making FAIR (Findable, Accessible, Interoperable, Reusable) research data and code available enables reuse and remixing of core research outputs, which can also lead to further citations for projects [25,26,27]. Published research proposals, protocols, and open notebooks act as advertisements for ongoing research and enable others to reuse methods, exposing the continuous and collaborative nature of scholarship.

To enable reuse, embrace open licenses. When it comes to innovative dissemination, the goal is usually that the materials are accessible to as large an audience as possible. If appropriate open licenses are not used, while materials may be free to access, they cannot be widely used, modified, or shared. The best in this case is the widely adopted Creative Commons licenses, CC BY or CC 0. Variations of these licenses are less permissive and can constrain reuse for commercial or derivative purposes. This limitation, however, prevents the use of materials in many forms of (open) educational resources and other open projects, including Wikipedia. Careful consideration should be given to licensing of materials, depending on what your intended outcomes from the project are (see Rule 1). Research institutes and funding bodies typically have a variety of policies and guidance about the use and licensing of such materials, and should be consulted prior to releasing any materials.

## Rule 5: Remix traditional outputs

Traditional research outputs like research articles and books can be complemented with innovative dissemination to boost impact; for example, by preparing accompanying nonspecialist summaries, press releases, blog posts, and visual/video abstracts to better reach your target audiences. Free media coverage can be an easy way to get results out to as many people as possible. There are countless media outlets interested in science-related stories. Most universities and large research organisations have an office for public affairs or communication: liaise with these experts to disseminate research findings widely through public media. Consider writing a press release for manuscripts that have been accepted for publication in journals or books and use sample forms and tools available online to assist you in the process. Some journals also have dedicated press teams that might be able to help you with this.

Another useful tool to disseminate traditional research outputs is to release a research summary document. This one- or two-page document clearly and concisely summarises the key conclusions from a research initiative. It can combine several studies by the same investigator or by a research group and should integrate two main components: key findings and fact sheets (preferably with graphical images to illustrate your point). This can be published on your institutional website as well as on research blogs, thematic hubs, or simply posted on your social media profiles. Other platforms such as ScienceOpen and Kudos allow authors to attach nonspecialist summaries to each of their research papers.

To maximise the impact of your conference presentations or posters, there are several steps that can be taken. For instance, you can upload your slides to a general-purpose repository such as Figshare or Zenodo and add a digital object identifier (DOI) to your presentation. This also makes it easier to integrate such outputs with other services like ORCID. You can also schedule tweets before and during any conferences, and use the conference hashtag to publicise your talk or poster. Finally, you can also add information about your contributions to email signatures or out-of-office messages [28].

## Rule 6: Go live

In-person dissemination does not just have to be at stuffy conferences. With research moving beyond the walls of universities, there are several types of places for more participatory events. Next to classic scientific conferences, different types of events addressing wider audiences have emerged. It is possible to hit the road and take part in science festivals, science slams, TEDx talks, or road shows.

Science slams are short talks in which researchers explain a scientific topic to a typically nonexpert audience. Similar to other short talk formats like TED talks, they lend themselves to being spread over YouTube and other video channels. A prominent example from the German-speaking area is Giulia Enders, who won the first prize in a science slam that took place in 2012 in Berlin. The YouTube video of her fascinating talk about the gut has received over 1 million views. After this success, she got an offer to write a book about the gut and the digestive system, which has since been published and translated into many languages. You never know how these small steps might end up having a wider impact on your research and career.

Another example is Science Shops, small entities which provide independent, participatory research support to civil society. While they are usually linked to universities, hacker and maker spaces tend to be community-run locations, where people with an interest in science, engineering, and art meet and collaborate on projects. Science festivals are community-based showcases of science and technology that take place over large areas for several days or weeks and directly involve researchers and practitioners in public outreach. Less formally, Science Cafés or similar events like Pint of Science are public engagement events in casual settings like pubs and coffeehouses.

Alternatively, for a more personal approach, consider reaching out to key stakeholders who might be affected by your research and requesting a meeting, or participating in relevant calls for policy consultations. Such an approach can be especially powerful in getting the message across to decision-makers and thought-leaders, although the resources required to schedule and potentially travel to such meetings means you should target such activities very carefully. And don't forget the value of serendipity—who knows who you'll meet in the course of your everyday meetings and travels. Always be prepared with a 30 second 'elevator pitch' that sums up your project in a confident and concise manner—such encounters may be the gateways to greater engagement or opportunities.

## Rule 7: Think visual

Dissemination of research is still largely ruled by the written or spoken word. However, there are many ways to introduce visual elements that can act as attractive means to help your audience understand and interpret your research. Disseminate findings through art or multimedia interpretations. Let your artistic side loose or use new visualisation techniques to produce intuitive, attractive data displays. Of course, not everyone is a trained artist, and this will be dependent on your personal skills.

Most obviously, this could take the form of data visualisation. Graphic representation of quantitative information reaches back to 'earliest map-making and visual depiction' [29]. As technologies have advanced, so have our means of visually representing data.

If your data visualisations could be considered too technical and not easily understandable by a nonexpert reader, consider creating an ad hoc image for this document; sometimes this can also take the form of a graphical abstract or infographic. Use online tools to upload a sample of your data and develop smart graphs and infographics (e.g., Infogr.am, Datawrapper, Easel.ly, or Venngage).

Science comics can be used, in the words of McDermott, Partridge, and Bromberg [30], to 'communicate difficult ideas efficiently, illuminate obscure concepts, and create a metaphor that can be much more memorable than a straightforward description of the concept itself'. McDermott and colleagues continue that comics can be used to punctuate or introduce papers or presentations and to capture and share the content of conference talks, and that some journals even have a 'cartoon' publication category. They advise that such content has a high chance of being 'virally' spread via social media.

As previously discussed, you may also consider creating a video abstract for a paper or project. However, as with all possible methods, it is worth considering the relative costs versus benefits of such an approach. Creating a high-quality video might have more impact than, say, a blog post but could be more costly to produce.

Projects have even successfully disseminated scientific findings through art. For example, The Civilians—a New York–based investigative theatre company—received a three-year grant to develop *The Great Immensity*, a play addressing the complexity of climate change. AstroDance tells the story of the search for gravitational waves through a combination of dance, multimedia, sound, and computer simulations. The annual Dance Your PhD contest, which began in 2007 and is sponsored by *Science* magazine, even asks scientists to interpret their PhD research as dance. This initiative receives approximately 50 submissions a year, demonstrating the popularity of novel forms of research dissemination.

## Rule 8: Respect diversity

The academic discourse on diversity has always included discussions on gender, ethnic and cultural backgrounds, digital literacy, and epistemic, ideological, or economic diversity. An approach that is often taken is to include as many diverse groups into research teams as possible; for example, more women, underrepresented minorities, or persons from developing countries. In terms of scientific communication, however, not only raising awareness about diversity issues but also increasing visibility of underrepresented minorities in research or including more women in science communication teams should be considered, and embedded in projects from the outset. Another important aspect is assessing how the communication messages are framed, and if the chosen format and content is appropriate to address and respect all audiences. Research should reach all who might be affected by it. Respect inclusion in scientific dissemination by creating messages that reflect and respect diversity regarding factors like gender, demography, and ability. Overcoming geographic barriers is also important, as well as the consideration of differences in time zones and the other commitments that participants might have. As part of this, it is a key responsibility to create a healthy and welcoming environment for participation. Having things such as a code of conduct, diversity statement, and contributing guidelines can really help provide this for projects.

The 2017 Progression Framework benchmarking report of the Scientific Council made several recommendations on how to make progress on diversity and inclusion in science: (1) A strategy and action plan for diversity should developed that requires action from all members included and (2) diversity should be included in a wide range of scientific activities, such as building diversity into prizes, awards, or creating guidance on building diversity and inclusion across a range of demographics groups into communications, and building diversity and inclusion into education and training.

## Rule 9: Find the right tools

Innovative dissemination practices often require different resources and skills than traditional dissemination methods. As a result of different skills and tools needed, there may be higher costs associated with some aspects of innovative dissemination. You can find tools via a more-complete range of sources, including the OpenUP Hub. The Hub lists a catalogue of innovative dissemination services, organised according to the following categories, with some suggested tools:

- Visualising data: tools to help create innovative visual representations of data (e.g., Nodegoat, DataHero, Plot.ly)
- Sharing notebooks, protocols, and workflows: ways to share outputs that document and share research processes, including notebooks, protocols, and workflows (e.g., HiveBench, Protocols.io, Open Notebook Science Network)
- Crowdsourcing and collaboration: platforms that help researchers and those outside academia to come together to perform research and share ideas (e.g., Thinklab, Linknovate, Just One Giant Lab)
- Profiles and networking: platforms to raise academic profile and find collaboration and funding opportunities with new partners (e.g., Humanities Commons, ORCID, ImpactStory)
- Organiding events: tools to help plan, facilitate, and publicise academic events (e.g., Open Conference Systems, Sched, ConfTool)
- Outreach to wider public: channels to help broadcast your research to audiences beyond academia, including policy makers, young people, industry, and broader society (e.g., Famelab, Kudos, Pint of Science)
- Publishing: platforms, tools, and services to help you publish your research (e.g., Open Science Framework, dokieli, ScienceMatters)
- Archive and share: preprint servers and repositories to help you archive and share your texts, data, software, posters, and more (e.g., BitBucket, GitHub, RunMyCode)

The Hub here represents just one attempt to create a registry of resources related to scholarly communication. A similar project is the 101 Innovations in Scholarly Communication project, which contains different tools and services for all parts of a generalised research workflow, including dissemination and outreach. This can be broadly broken down into services for communication through social media (e.g., Twitter), as well as those designed for sharing of scholarly outputs, including posters and presentations (e.g., Zenodo or Figshare). The Open Science MOOC has also curated a list of resources for its module on Public Engagement with Science, and includes key research articles, organisations, and services to help with wider scientific engagement.

## Rule 10: Evaluate, evaluate, evaluate

Assess your dissemination activities. Are they having the right impact? If not, why not? Evaluation of dissemination efforts is an essential part of the process. In order to know what worked and which strategies did not generate the desired outcomes, all the research activities should be rigorously assessed. Such evaluation should be measured via the use of a combination of quantitative and qualitative indicators (which should be already foreseen in the planning stage of dissemination; see Rule 1). Questionnaires, interviews, observations, and assessments could also be used to measure the impact. Assessing and identifying the most successful practices will give you the evidence for the most effective strategies to reach your audience. In addition, the evaluation can help you plan your further budget and minimise the spending and dedicating efforts on ineffective dissemination methods.

Some examples of quantitative indicators include the following:

- Citations of publications;
- alternative metrics related to websites and social media platforms (updates, visits, interactions, likes, and reposts);
- numbers of events held for specific audiences;
- numbers of participants in those events;
- production and circulation of printed materials;
- media coverage (articles in specialised press newsletters, press releases, interviews, etc.); and
- how much time and effort were spent on activities.

Some examples of qualitative indicators include the following:

- Visibility in the social media and attractiveness of website;
- newly established contacts with networks and partners and the outcomes of these contacts;
- feedback from the target groups; and
- share feedback within your group on what dissemination strategies seemed to be the most effective in conveying your messages and reaching your target audiences.

We recognise that researchers are usually already very busy, and we do not seek to pressurise them further by increasing their burdens. Our recommendations, however, come at a time when there are shifting norms in how researchers are expected to engage with society through new technologies. Researchers are now often partially evaluated based on such, or expected to include dissemination plans in grant applications. We also do not want to encourage the further fragmentation of scholarship across different platforms and 'silos', and therefore we strongly encourage researchers to be highly strategic in how they engage with different methods of innovative dissemination. We hope that these simple rules provide guidance for researchers and their future projects, especially as the tools and services available evolve through time. Some of these suggestions or platforms might not work across all project types, and it is important for researchers to find which methods work best for them.

## Acknowledgments

Many thanks to everyone who engaged with the workshops we conducted as part of this grant award.

## References

1. Jarreau P. All the Science That Is Fit to Blog: An Analysis of Science Blogging Practices. *LSU Doctoral Dissertations*, January 2015 [cited 2020 Mar 23]. Available from: https://digitalcommons.lsu.edu/gradschool_dissertations/1051
2. Bik HM, Goldstein MC. An Introduction to Social Media for Scientists. PLoS Biol 11(4):e1001535. 2013. https://doi.org/10.1371/journal.pbio.1001535 PMID: 23630451
3. Yammine SZ, Liu C, Jarreau PB, Coe IR. Social Media for Social Change in Science. *Science* 360 (6385): 162–63. https://doi.org/10.1126/science.aat7303. 2018.
4. Bradley JC. Open Notebook Science using blogs and wikis. Nature Precedings, https://doi.org/10.1038/npre.2007.39.1 2007.
5. Clinio A, Albagli S. Open Notebook Science as an Emerging Epistemic Culture within the Open Science Movement. *Revue Française Des Sciences de l'information et de La Communication*, no. 11 (August). https://doi.org/10.4000/rfsic.3186. 2017.

6. Harding RJ. Open Notebook Science Can Maximize Impact for Rare Disease Projects. PLoS Biol 17 (1):e3000120. 2019. https://doi.org/10.1371/journal.pbio.3000120 PMID: 30689629

7. Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, et al. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59(11):977–984. 2009.

8. Williams AJ, Peck L, Ekins S. The new alchemy: Online networking, data sharing and research activity distribution tools for scientists [version 1; peer review: 2 approved, 1 approved with reservations]. *F1000Research*, 6:1315. 2017. https://doi.org/10.12688/f1000research.12185.1 PMID: 28928951

9. Stilgoe J, Lock SJ, Wilsdon J. Why Should We Promote Public Engagement with Science? *Public Understanding of Science* 23(1):4–15. 2014. https://doi.org/10.1177/0963662513518154 PMID: 24434705

10. Sengupta P, Shanahan MC. Open Science, Public Engagement and the University. ArXiv:1702.04855 [Physics]. 2017 [cited 2020 Mar 23]. Available from: http://arxiv.org/abs/1702.04855

11. Alperin JP, Muñoz Nieves C, Schimanski L, Fischman GE, Niles MT, McKiernan EC. How Significant Are the Public Dimensions of Faculty Work in Review, Promotion, and Tenure Documents?, eLife; 8: e42254 2019. https://doi.org/10.7554/eLife.42254 PMID: 30747708

12. Wilson PM, Petticrew M, Calnan MW, Nazareth I. Disseminating Research Findings: What Should Researchers Do? A Systematic Scoping Review of Conceptual Frameworks. *Implementation Science* 5(1):91. https://doi.org/10.1186/1748-5908-5-91. 2010.

13. Ekins S, Perlstein EO. Ten Simple Rules of Live Tweeting at Scientific Conferences. PLoS Comput Biol 10(8):e1003789. 2014. https://doi.org/10.1371/journal.pcbi.1003789 PMID: 25144683

14. Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. Ten Simple Rules for Effective Online Outreach. PLoS Comput Biol 11(4):e1003906. 2015. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439

15. Blackmore P, Kandiko CB. Motivation in Academic Life: A Prestige Economy. *Research in Post-Compulsory Education* 16 (4): 399–411. https://doi.org/10.1080/13596748.2011.626971. 2011.

16. Kraker P, Bachleitner R, Banelytė V, Hasani-Mavriqi I, Luzi D, Ruggieri R, et al. Deliverable D4.1 –Practices Evaluation and Mapping: Methods, Tools and User Needs. OpenUP project report. https://doi.org/10.5281/zenodo.2557357. 2019.

17. Bornmann L. Measuring the Societal Impact of Research. *EMBO Reports* 13 (8): 673–76. 2012. https://doi.org/10.1038/embor.2012.99 PMID: 22777497

18. Vicente-Saez R, Martinez-Fuentes C. Open Science Now: A Systematic Literature Review for an Integrated Definition. *Journal of Business Research* 88 (July): 428–36. https://doi.org/10.1016/j.jbusres.2017.12.043. 2018.

19. Pontika N, Knoth P, Cancellieri M, Pearce S. Fostering Open Science to Research Using a Taxonomy and an ELearning Portal. In: iKnow: 15th International Conference on Knowledge Technologies and Data Driven Business, 21–22 Oct 2015, Graz, Austria [cited 2020 Mar 23]. Available from: http://oro.open.ac.uk/44719/

20. McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. How Open Science Helps Researchers Succeed. *ELife* 5 (July). https://doi.org/10/gbqsng. 2016.

21. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink CHJ. The Academic, Economic and Societal Impacts of Open Access: An Evidence-Based Review. *F1000Research* 5 (September): 632. https://doi.org/10.12688/f1000research.8460.3. 2016.

22. Shuai X, Pepe A, Bollen J. How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. PLoS ONE 7 (11): e47523. 2012. https://doi.org/10.1371/journal.pone.0047523 PMID: 23133597

23. Powell K. Does It Take Too Long to Publish Research? *Nature News* 530 (7589): 148. https://doi.org/10.1038/530148a. 2016.

24. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Nathalie Percie du Sert N, et al. A Manifesto for Reproducible Science. *Nature Human Behaviour* 1(1): 0021. https://doi.org/10.1038/s41562-016-0021. 2017.

25. Drachen T, Ellegaard O, Larsen A, Dorch S. Sharing Data Increases Citations. *LIBER Quarterly* 26(2): 67–82. https://doi.org/10.18352/lq.10149. 2016.

26. Pasquetto I, Randles B, Borgman C. On the use of Scientific Data. *Data Science Journal* 16(0). https://doi.org/10.5334/dsj-2017-008. 2017.

27. Piwowar HA, Day RS, and Fridsma DB. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." PLoS ONE 2 (3): e308. 2007. https://doi.org/10.1371/journal.pone.0000308 PMID: 17375194

**28.** Teperek M. How to Make the Most of an Academic Conference–a Checklist for before, during and after the Meeting." *Impact of Social Sciences blog* [Internet]. March 16, 2018 [cited 2020 Mar 23]. Available from: http://blogs.lse.ac.uk/impactofsocialsciences/2018/03/16/how-to-make-the-most-of-an-academic-conference-a-checklist-for-before-during-and-after-the-meeting/

**29.** Friendly M. A Brief History of Data Visualization. In: *Handbook of Data Visualization*. Chen C, Härdle W, Unwin A, editors. Springer Handbooks Comp.Statistics. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-33037-0_2. 2008. Pp.15–56.

**30.** McDermott JE, Partridge M, Bromberg Y. 2018. "Ten Simple Rules for Drawing Scientific Comics." PLoS Comput Biol 14(1):e1005845. https://doi.org/10.1371/journal.pcbi.1005845 PMID: 29300733

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for writing algorithmic bioinformatics conference papers

**Paul Medvedev**[1,2,3]*

**1** Department of Computer Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania, United States of America, **2** Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania, United States of America, **3** Center for Computational Biology and Bioinformatics, The Pennsylvania State University, University Park, Pennsylvania, United States of America

* pzm11@psu.edu

## Author summary

Conferences are great venues for disseminating algorithmic bioinformatics results, but they unfortunately do not offer an opportunity to make major revisions in the way that journals do. As a result, it is not possible for authors to fix mistakes that might be easily correctable but nevertheless can cause the paper to be rejected. As a reviewer, I wish that I had the opportunity to tell the authors, "Hey, you forgot to do this really important thing, without which it is hard to accept the paper, but if you could go back and fix it, you might have a great paper for the conference." This lack of a back and forth can be especially problematic for first-time submitters or those from outside the field, e.g., biologists. In this article, I outline Ten Simple Rules to follow when writing an algorithmic bioinformatics conference paper to avoid having it rejected.



**Citation:** Medvedev P (2020) Ten Simple Rules for writing algorithmic bioinformatics conference papers. PLoS Comput Biol 16(4): e1007742. https://doi.org/10.1371/journal.pcbi.1007742

**Editor:** Russell Schwartz, Carnegie Mellon University, UNITED STATES

**Published:** April 2, 2020



**Data Availability Statement:** There is no data.

**Funding:** This work has been supported in part by the National Science Foundation (www.nsf.gov) awards DBI-1356529 and CCF-1439057 to PM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

As a frequent program committee (PC) member of bioinformatics conferences, I sometimes find it frustrating to see a paper that potentially has a great contribution be rejected because of the way it was written. I wish that I had the opportunity to tell the authors, "Hey, you forgot to do this really important thing, without which it is hard to accept the paper, but if you could go back and fix it, you might have a great paper for the conference." In our conference format, this type of back and forth is usually not possible. This motivated this article so that newcomers to the field have a chance to know in advance what a potential reviewer might look for in an algorithmic bioinformatics conference paper.

What do I mean by algorithmic bioinformatics conference paper? I am thinking of a subset of papers submitted to the following conferences:

- International Conference on Research in Computational Molecular Biology (RECOMB)
- Workshop on Algorithms in Bioinformatics (WABI)
- The proceedings track of Intelligent Systems for Molecular Biology (ISMB). Note that ISMB is very diverse, with different types of tracks, and this article only refers to those focusing on methods. Certain tracks may also feature papers focusing on biology, in which case some of the rules below may not apply.
- RECOMB Satellite Workshop on Massively Parallel Sequencing (RECOMB-Seq)

Other conference likely fall into this category, although I am not as personally familiar with them. The types of papers I am thinking of is the subset of papers that take an algorithm-based approach to solving a bioinformatics problem. This is largely intended to contrast with papers more rooted in statistical methodology, in which the standards are a bit different. I also focus on conference reviews, in which the process is a bit different than for a journal. When reviewing a paper for a bioinformatics journal like Oxford's *Bioinformatics*, there is of course an opportunity for the authors to address any limitations in a revision.

I want to also add a disclaimer that this is not in any way an official statement about what PC members would look for in a review. Each conference has an official call for papers. It may sometimes state or imply necessary elements for submission, although these are typically at a fairly high level. As far as I know, there is no official policy at the level of detail presented here, and the things that each PC member looks for do not completely overlap. There is a diversity of standards and that is why each paper has multiple PC members reviewing it.

## The rules

### Rule 1: Make sure to clearly and succinctly state what the main novel contribution of the paper is

When I review, the first thing I try to identify is as follows: What is the main novel contribution of the paper? Is it an idea, a theorem, an algorithm, or a tool (i.e., software) people can use? Sometimes a paper has all these components, but not all of them contribute to the novelty of the paper. Here are some examples:

1. The paper contains an algorithm and a tool that implements the algorithm. The algorithm itself may be a simple modification of what is previously known, but the algorithm is implemented in a novel software tool for an important biological problem. If the tool performance is an improvement over previous tools, then the tool is the main contribution.

2. In another example, the main novelty is in the algorithm or in its analysis, and this is what the reader is intended to take away from the paper. The paper may have implemented a tool, but the intention of the tool is to only be a prototype to test the feasibility of the idea. The tool is not the main contribution.

3. Sometimes, the main contribution of the paper is novel biological findings without any methodological (either algorithmic or software) novelty. This is not really within the scope of an algorithmic bioinformatics conference, which has to be methodological. Certainly, having novel biological findings can serve to demonstrate the strength of the methodological contribution. But if you discover a cure for cancer by applying existing software, then it is probably outside the scope.

It is up to the authors to make the main contribution of the paper crystal clear to the reader. As a reviewer, I will then base my evaluation on what the authors claim. For example, as I will describe in Rule 8, the standards for evaluating a paper whose main contribution is a tool are different for a paper whose main contribution is a theorem. If the authors' claim is not clearly stated, then I will do my best to guess what it is. But if I make a mistake, then I may end up evaluating the paper from a completely incorrect angle.

### Rule 2: Give context within prior algorithmic work

A common scenario in which this rule is not followed is when the authors developed a method for a particular biological data set and there are no other tools designed specifically for this kind of data set or problem. However, the problem and/or solution might be very similar to what has been previously studied. For instance, many problems come down to clustering of some data points (e.g., genes in a network or reads from a sequencing experiment) or to some version of sequence alignment. The algorithmic context of such a paper is, at least in part, clustering or, respectively, alignment algorithms. Sometimes the authors provide the biological context (e.g., what is the relationship to previous approaches to finding genes in a network?) but leave out the algorithmic one (e.g., what is the relationship to previous clustering algorithms). Why is this particular problem or data set different enough so that standard clustering or alignment techniques do not apply? If the authors present a clustering algorithm for the problem but do not answer this question in the introduction, then their contribution is not placed in the algorithmic context—which makes it hard to evaluate its novelty.

## Rule 3: Make the writing clear

Some papers will contain many spelling and grammatical mistakes or ambiguous notation and terminology. These of course should be avoided, and at least spelling can be easily improved by using a spell checker. I try to do the best I can to understand the contribution of the paper, and often I do understand it in spite of these problems. In such cases, it does not greatly influence my overall decision about the paper, and I generally trust the authors to clean up the paper before publication (if it is accepted). In other cases, I cannot understand the paper after a reasonable amount of time trying. This is especially the case with ambiguous notation or terminology. In these cases, I simply cannot evaluate the paper's contribution.

## Rule 4: Do not write the paper in the style of a biology journal

In biology journals, the methods section is often written as a step-by-step manual necessary to reproduce the results (i.e., a pipeline of processing steps on the data). This type of presentation focuses on implementation details and reproducibility rather than highlighting the novelty of the algorithm. Even if the method is novel, when it is written in this style it is hard for the reader to identify and understand the novel parts. Another aspect of this is that for a biology journal, the results section comes before the methods section. Doing this for an algorithmic bioinformatics paper is not in it of itself a problem, but it usually correlates with not enough focus being given to the method.

## Rule 5: Make sure that claims in the introduction are supported by the rest of the paper

For example, the authors claim that their tool is the fastest to date for a problem, but the results section only contains a comparison against one other tool or only on a narrow type of data. In such cases, I simply ask the authors to tone down their claims. However, sometimes the claims are central to the claimed importance of the paper, in which case the lack of proper evaluation feels a bit disingenuous. Another example is the bait and switch, in which the introduction claims that the paper presents an algorithm for some interesting problem, but what ends up being evaluated in the results is an algorithm for a slightly different problem.

## Rule 6: Make sure there is either a strong theoretical contribution or an experimental evaluation

Some contributions are theoretical—a powerful idea, a way of thinking about a problem, or a theorem that can be applied by other algorithm developers. These papers require a lot of work on the modeling or theoretical side, and it can be justifiable if experimental results are either not included or limited. However, in most other cases, experimental evaluation is essential to a paper. If this is missing or is inappropriate to the problem, it can be impossible to evaluate the strength of the contribution.

## Rule 7: Compare against other work

The authors sometimes find it obvious that their method should work much better than anything else out there. They may be right, but it is important to demonstrate this in the paper by finding the most compelling alternative approach and comparing their method against it.

When doing an empirical comparison, the authors have wide leeway in choosing which data sets, computing configuration, or parameters to use. This is sometimes referred to as researcher degrees of freedom [1]. It is important that the authors are forthright about how their choices affect the evaluation. For example, while it is normal to use data sets that would demonstrate the advantages of the presented algorithm, it should nevertheless be made clear that the data sets were chosen with this in mind and that there may be other data sets on which the other tools would perform better.

## Rule 8: If the main contribution of the paper is a tool, then the software should be usable

At the very least, I should be able to download the software, install it, and run it on a toy input that is provided in the download. If I can see that the tool already has some users (e.g., through GitHub activity), then this is enough to demonstrate its usability, and I may not bother to try it out myself. On the other hand, if the paper contains a tool that is only a prototype and is not the main contribution, then the usability of the software is not something I consider important. However, I still expect it to at least be publicly available for download.

## Rule 9: Give a precise description of the algorithm, argue its correctness, and verify the correctness of the method's computations explicitly in the experiments

This rule is especially applicable when the main contribution of the paper is an algorithm. It has three distinct parts. The first is to describe the algorithm precisely. This means to explicitly state (1) what the input is and what the assumptions are made about it are, (2) what the output is, independently of what the algorithm is, and (3) what the algorithm that converts the input to the output is. These should be stated in a way that is unambiguous, using mathematical notation and/or pseudocode to the extent it facilitates preciseness.

Second, it should be argued why the algorithm achieves its stated goal. Ideally, the goal should be stated explicitly as a problem formulation [2], in which case a formal proof of correctness (or at least a coherent argument) should be given. Sometimes, the correctness is obvious from the construction, especially in the case of a data structure, and a separate proof is not needed. It should also be made clear if the algorithm solves the problem exactly or is a heuristic. If the algorithm is a heuristic and no argument of correctness is necessary, it should be made clear.

Finally, if the algorithm is evaluated empirically, it is essential that the correctness of the algorithm is explicitly verified for the experiments, if possible. This can be a simple one line that says, for example, "We verified that the new data structure gives the same answers to queries as the previous one on all the evaluated data sets." However, without this check, how does the reader know that the algorithm is not twice as fast as the competition just because it has a bug? Sometimes, however, a verification is not possible when the ground truth is ambiguous.

## Rule 10: Give a theoretical and/or empirical analysis of running time or memory usage

In most cases, it is important for an algorithmic bioinformatics paper to present the running time and memory usage of the algorithm, either through experimental evaluation and/or theoretical analysis. This is a very natural thing to do for computer scientists, but I sometimes find that researchers with a different background forget to include this. In other cases, the authors do not include any memory or time analysis because they know that it is tiny and besides the main point, but it may not be at all obvious to the reader. In such cases, a simple statement to the effect that the memory usage or running time is negligible and/or unimportant would suffice. Finally, make sure to state the specific details of the machine used, to the extent they are relevant to your algorithm; for example, "Intel Xeon CPU with 512 GB of RAM and 64 cores at 2.10 GHz."

## Conclusion

This list includes only the most basic rules and is not intended to be exhaustive. In a competitive venue, a paper is usually accepted based on its strengths rather than a lack of weaknesses, and following these simple rules will not necessarily get your paper accepted. However, in my experience, breaking these rules can significantly decrease a paper's chance to be accepted.

These rules are also focused on aspects that are somewhat specific to algorithmic bioinformatics in relation to biology. There are of course much broader aspects about how to write a scientific paper, e.g., cohesion, conciseness, clarity, structure, and argumentation. These are outside the scope of this paper, but there are excellent books [3, 4] and even other Ten Simple Rules articles [5] that address these aspects.

## References

1. Simmons JP, Nelson LD, Simonsohn U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychological Science. 2011; 22 (11):1359–1366. https://doi.org/10.1177/0956797611417632 PMID: 22006061
2. Medvedev P. Modeling biological problems in computer science: a case study in genome assembly. Briefings in bioinformatics. 2019; 20(4):1376–1383. https://doi.org/10.1093/bib/bby003 PMID: 29394324
3. Schimel J. Writing science: how to write papers that get cited and proposals that get funded. OUP USA; 2012.
4. Williams J. Toward Clarity and Grace; 1990.
5. Mensh B, Kording K. Ten simple rules for structuring papers. PLoS Comput Biol. 2017; 13(9).

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for getting started on Twitter as a scientist

**Veronika Cheplygina[1], Felienne Hermans[2,3]*, Casper Albers[4], Natalia Bielczyk[5], Ionica Smeets[6]**

**1** Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, **2** Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands, **3** Software Engineering Research Group, Delft University of Technology, Delft, The Netherlands, **4** Heymans Institute for Psychological Research, University of Groningen, Groningen, The Netherlands, **5** Stichting Solaris Onderzoek en Ontwikkeling, Nijmegen, the Netherlands, **6** Science Communication and Society, Institute of Biology, Leiden University, Leiden, The Netherlands

* felienne@gmail.com



OPEN ACCESS

**Citation:** Cheplygina V, Hermans F, Albers C, Bielczyk N, Smeets I (2020) Ten simple rules for getting started on Twitter as a scientist. PLoS Comput Biol 16(2): e1007513. https://doi.org/10.1371/journal.pcbi.1007513

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** February 10, 2020



**Funding:** The authors received no specific funding for this article.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Twitter is one of the most popular social media platforms, with over 320 million active users as of February 2019. Twitter users can enjoy free content delivered by other users whom they actively decide to follow. However, unlike in other areas where Twitter is used passively (e.g., to follow influential figures and/or information agencies), in science it can be used in a much more active, collaborative way: to ask for advice, to form new bonds and scientific collaborations, to announce jobs and find employees, to find new mentors and jobs. This is particularly important in the early stages of a scientific career, during which lack of collaboration or delayed access to information can have the most impact.

For these reasons, using Twitter appropriately [1] can be more than just a social media activity; it can be a real career incubator in which researchers can develop their professional circles, launch new research projects and get helped by the community at various stages of the projects. Twitter is a tool that facilitates decentralization in science; you are able to present yourself to the community, to develop your personal brand, to set up a dialogue with people inside and outside your research field and to create or join professional environment in your field without mediators such as your direct boss.

This article is written by a group of researchers who have a strong feeling that they have personally benefited from using Twitter, both research-wise and network-wise. We (@DrVeronikaCH, @Felienne, @CaAl, @nbielczyk_neuro, @ionicasmeets) share our personal experience and advice in the form of ten simple rules, and we hope that this material will help a number of researchers who are planning to start their journey on Twitter to take their first steps and advance their careers using Twitter.

## Twitter terms

Before we start with the rules, inspired by [2] we briefly introduce a number of relevant Twitter terms.

- Timeline—The tweets from the people you follow.
- Retweet (RT)—sharing a tweet that was originally made by someone else.
- Quote-tweet—sharing a tweet by someone else in a quote, while adding your own comments.

- Like (), used for showing you like a tweet—a fast way to give feedback without replying. There is no similar function for disliking a tweet.
- Notifications—Tweets that mention you and replies, retweets and likes for your tweets.
- Mentioning (@)—If you mention someone with their handle ("This paper by @CaAl is great"), your tweet will show up in their notifications.
- Direct message (DM)—A private message that is only visible to the sender and the specifically identified recipients. By default, regular Twitter messages are visible to the whole world, including (via search engines such as Google) people who do not have a Twitter account.
- Hashtag (#)—used to make it easier to find tweets with a common theme by defining ad hoc keywords, for instance tweets about a conference (#ICA19) or career talks (#PhDChat).
- List—a list of Twitter users that can be public (followed by anyone) or private. Lists can be used to follow accounts that tweet about specific topics, but which you don't want to follow yourself.
- Hat Tip or Heard Through (HT), used for thanking the source of a tweet.
- Subtweeting—Tweeting about somebody without explicitly mentioning their handle, so that they are not informed of your comment (see 'Mentioning').
- Live-tweeting—Tweeting short summaries of an event, for example of a conference talk, as it is happening.
- Thread—A series of tweets on one subject, for instance ten tweets about a new research paper.

## Rule 1: Start somewhere, but show up

To get into the habit of using Twitter, you need to do just that: start. You can do this even before you finish reading this article! You don't need to know everything about Twitter before posting a tweet, just like you don't need to know everything about running a marathon before going for a jog. Create an account (you can change your username later) and tweet something—for example, a comment about a paper you are writing or a conference you are going to. This way you can use ideas that you already have—but put them into the form of a tweet, which aligns with the advice of not treating research and outreach as separate entities [1].

A bigger challenge is to continue tweeting regularly once you have started, which is essential to habit change [3], similar to going running every few days. There is no one-size-fits-all set of rules for this, but we provide a few suggestions here. Once that paper is completed or the conference is over, perhaps you can follow other people in your field and respond to their tweets first, or check hashtags where everybody is invited to contribute, such as #PhDChat or #ECRChat. Such general hashtags are also great ways to discover and follow people who you would not run into at your department or conferences. Hashtags are good places to start posting your own questions or content—for an easy start, try #AcademicsWithCats, where you can show off your cat helping you with reviewing duties (Fig 1).

## Rule 2: Discover opportunities in academia

To Early Career Researchers (ECRs), Twitter has become an invaluable source of information [4,5]. One can follow granting agencies, particular labs and dedicated career columns in

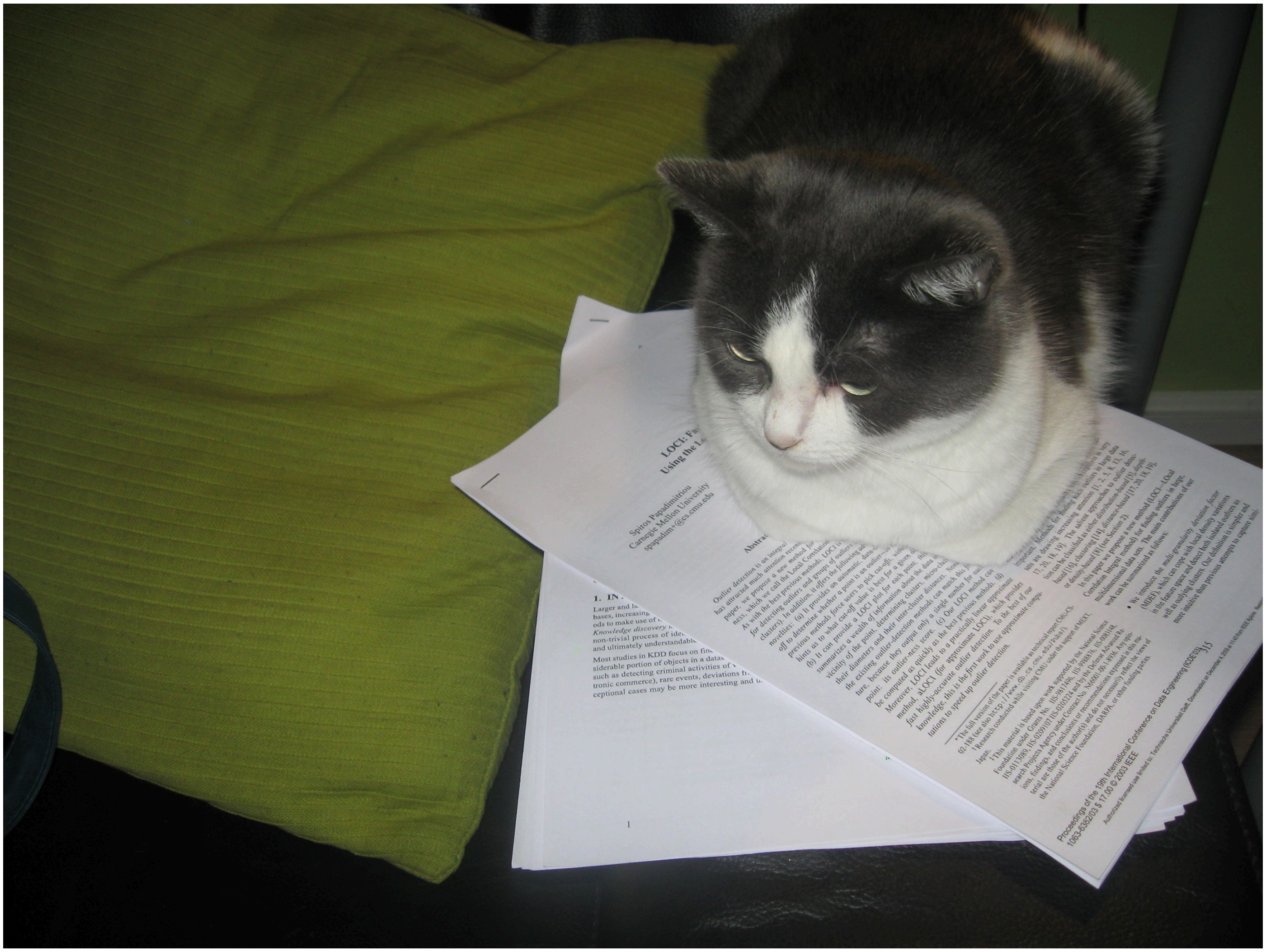


**Fig 1. Cats.** Pictures that someone might post on Twitter to convey they are reading papers AND like cats. Possible relevant hashtags include #AcademicsWithCats, #Caturday, #ECRchat (chat as in conversation, not as in cat in French).

https://doi.org/10.1371/journal.pcbi.1007513.g001

popular research journals, which allows for tracking the first-hand information on recently opened positions, grant calls, and new trends on the academic job market.

But even more importantly, Twitter culture has exposed a part of academia that has traditionally always been hidden from view, namely the inception of new research activities. Now, ECRs can observe and even join the process of creating national or international research projects (for instance, [6] stems from a discussion at Twitter and Bik et al [1] write their work resulted from online interactions). Senior researchers openly share ideas through Twitter and this can lead to the development of new concepts which often move on to become fully-fledged research projects.

Many researchers active on Twitter are also open about their everyday struggles, share their frustrations from rejections and bureaucracy, and support each other emotionally [7]. This is especially valuable for ECRs who have experienced the black side of academic life: patronizing behaviors, inequality and lack of inclusivity in research projects. For these researchers, Twitter gives an opportunity to follow senior role models who have managed to develop successful research careers and are willing to openly share their difficult experiences from the past. Twitter can play the role of a distributed mentoring system which can have a profound impact on a researcher's self-image and can help to mitigate imposter syndrome. Similar reasons are listed by [8], who crowdsourced over 400 responses from academics on why they use Twitter.

For ECRs, starting Twitter activity may be hard. Thus, we recommend joining a peer group, together with members of your local research group, together with other collaborators or friends in the research community. In this way, you can make sure that you have a few followers right at the start, and that your peer group helps you in distributing your content.

## Rule 3: Tweet stuff

Twitter is useful for absorbing and sending information, but the potential for interaction with fellow scientists is where its real power lies [9]. Especially for new Twitter users, it is good practice to retweet other users' content. Retweeting means sharing someone else's tweet with your followers [10]. This way, your timeline demonstrates the scientific topics of your interest. You can either retweet directly or quote-tweet, which means adding your own commentary to the tweet. Quote-tweeting adds a personal flavor to your tweets.

Interact with others by asking them their opinion on your (re)tweet. Interactivity can be increased by using (simple) polls. Do not hesitate to ask scientists questions about their work (e.g., "can you share the code for that paper?"): they signed up to Twitter because they intend to interact. Also, feel free to ask questions, both content-related ("Which *R* package can we use for this?") and logistic ("How do you handle requests for letters of recommendation?").

Many scientists are bilingual on Twitter, tweeting both in their mother tongue and in English. Do not worry that this might be confusing: Twitter users can easily skip over tweets in languages they are not proficient in, or make use of the automatic translation facility that is offered on every tweet in a foreign language.

## Rule 4: Learn the rules

It should go without saying, but as with any situation where you communicate with others, you should treat them with respect. If you wouldn't do something in real life, you probably should not do it on Twitter. Next to conventional politeness, there is also some basic Twitter-etiquette that you should follow: credit ideas from others by mentioning them with HT (there is some discussion whether HT means "hat tip" or "heard through"—but there is absolutely no discussion about whether to use it) and their username. Be careful about subtweeting people (i.e., talking about them or their work without mentioning them explicitly in your tweet). It is like gossiping, but then clearly visible to everyone. This is especially problematic if you are in a position of privilege with respect to the person you are quoting. Repeatedly tweeting at somebody who is not interested is another behavior to avoid. This is called "sealioning", and is one of the negative sides of Twitter experienced at a higher rate by underrepresented groups on Twitter.

Use hashtags when appropriate. You can give tips for people who are interesting to follow on Friday with: #FollowFriday or share stories of colleagues who are helping others with #AcademicKindness. Next to hashtags, there are many memes that are like little inside jokes; use those only when you really understand them. One of our favorite examples is "Asking for a friend" used in tweets where you are asking something for yourself that is a bit embarrassing. One example is: "Why is Reviewer 2 always the one who says you should have written on a different topic using a different method? Asking for a friend" [11].

Also, do not respond impulsively if someone is critical towards your research on Twitter. Diplomacy is one of the key components to building a scientific reputation [12].

It might be good to check whether your institute has any social media guidelines (and stick to those). For example, Stanford University tells researchers not to endorse commercial partners (https://ucomm.stanford.edu/policies/social-media-guidelines/) and the University of Oxford warns employees to be mindful of what pictures might reveal in the background (https://www.admin.ox.ac.uk/personnel/during/socialmedia/). Some employers, for example Leiden University (https://www.medewerkers.universiteitleiden.nl/binaries/content/assets/ul2staff/reglementen/communicatie-en-marketing/gedragscode-social-media-2012.pdf) may ask you to add a disclaimer that opinions are your own.

## Rule 5: Take care of yourself

Unfortunately, online conversation can easily go off the rails and you might need to protect yourself from trolls and nasty discussions. Curate who you choose to follow and be prepared to mute or block people. Muting people means that you will generally not see their tweets anymore, but they will still be able to read your tweets and reply to them. You can also mute specific words for which you do not want to see tweets: names of politicians, TV shows, and many other topics. This can be useful if you want to follow a colleague who is both tweeting about interesting research and ranting about politics. When muting somebody, the person will not be aware that you muting them. Blocking is a more drastic measure than muting and makes the blocked user unable to read your tweets or react to them. The blocked user will be aware that they are blocked, if they attempt to read your tweets. There is no single guideline on when to block somebody—some may only support blocking trolls, and refrain from blocking other scientists who may be critical, while others believe it's perfectly fine to mute or block any account [13]. A good rule of thumb might be to block somebody if they have a repeated negative effect on your Twitter experience. For example a critic with many followers who frequently quote-tweets your posts might want to start a discussion, but could end up unleashing many trolls, in which case blocking would be the only sure way to prevent the same from happening again.

Never feel pressured to read everything on Twitter. It is perfectly okay to <u>not</u> look at Twitter for hours or even days. You definitely don't need to read up your entire timeline when you have been away, just join the conversation at the point where you fall in again. See it as dipping into the stream. It might be good to read up on your mentions (rather than all notifications), since these are personal remarks and questions just for you.

In fact, using Twitter irresponsibly can yield negative impact on mental health. One should remember that tweeting can be addictive and should therefore be self-managed with care. Getting notifications on social media activates the same dopamine loop as gambling [14]. When you become addicted to Twitter and you are not able to check your phone, adrenaline and cortisol are released to your body [15]. Therefore, it is beneficial to dose the amount of Twitter in your daily life: set fixed times during the day or during the week when you remain logged into your Twitter account, and switch it off otherwise. One strategy is to limit Twitter use to your time on public transport, or while in the queue to order a coffee.

## Rule 6: Build your own community

A nice feature of Twitter is its asymmetry: you do not have to follow everyone who follows you and vice versa. You can follow big names you have never met and should not be afraid to join a conversation with more senior researchers or people you do not know. Following diverse voices will broaden your horizon.

To a certain extent, you can also determine the kind of community your followers form. Twitter offers the possibility for communicating directly with people who would otherwise not find out about your work and field. This is a very relevant form of science communication, since it is a transaction of ideas and not just a transmission [16]. If this appeals to you, you could share and explain papers that might be interesting to people outside your field, comment on news items from your expertise and answer questions from people you do not know. And vice versa: posing questions is also a great way to build a community, since people who share common interests will find each other on your timeline.

## Rule 7: Interface with real life

Twitter is a great way to make networking more easy and fun and less scary. For example, it can change your experience in attending a conference [2]. Firstly, you can find interesting people to follow by browsing through the list of users who include the conference hashtag in their tweets. If a conference does not have a well-defined hashtag, try a few options to see what most people are using (for example #sigcse19 versus #sigcse2019; [2]). One way to use the conference hashtag is to announce your presence or your research ahead of time. Another possibility is to live-tweet—briefly summarize talks you are attending—which is helpful both for other people at the conference, as well as people who could not travel, thus potentially communicating the content to a more diverse audience. It also helps if your Twitter avatar includes your face, so that people who also follow the conference hashtag can spot you in the coffee breaks at the conference site.

Following people from conferences has value *after* the conference too. Twitter is a good way to stay in touch after a conference, aa it provides a low-cost way of tracking what people are working on or interested in. Knowing a bit about their interests in research and in life will make it a lot easier and less scary to say "hi" to people the year after!

This advice also holds outside of conferences: if you are approaching people you do not know, for example, to ask for a committee membership or a collaboration—you can check out their Twitter account and get an idea, not just of what they work on, but also their style of communication in general. Finally, when you are visiting a conference or university, you can advertise your visit in advance and perhaps plan extra meetings or talks.

There are also possibilities to make your Twitter network more extensive and more interesting by interfacing with real life. Put your Twitter handle on slides (see Fig 2) and other material, so people that you meet at conferences or talks can follow you, thus further extending your Twitter network.

## Rule 8: Spread your message

Whenever you have some scientific accomplishment, you can share this information by, for instance, sharing a link to a preprint or vacancy notice. In order to get the best exposure, summarize the content and include an image and an appropriate hashtag (e.g., #phdposition). You can "tag" (i.e., mention the usernames of) scientists if you think your tweet is of special interest to them, but do not overdo this. Retweeting can also be a way to put scientists whom you feel deserve more attention in the spotlight.

Adding humor to your tweets (e.g., [18, 19]) can help to get you noticed by more users who might then follow you and also receive updates of your more serious contributions. Twitter is not only useful to communicate with fellow scientists, but also with people outside of science [20]. This does, however, require a different style of communication [21].

Although Twitter is accessible worldwide, it is useful to consider the time of tweeting things, especially when your tweet has more of a local audience. Many people do not read the whole of their timeline and so may never see tweets that were sent while they were asleep. Also keep in mind that not all scientists work at weekends, so science Twitter is a bit quieter then than on

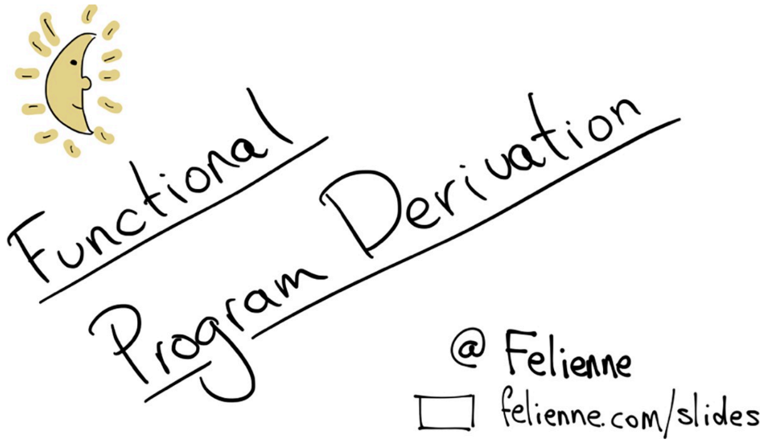



**Fig 2. Title slide.** Example of adding your Twitter handle to your slides. Source: [17].

https://doi.org/10.1371/journal.pcbi.1007513.g002

weekdays. Tweets for global audiences can be retweeted at a later moment in order to reach different time zones.

## Rule 9: Be a real person

Even if you use Twitter only for professional purposes, consider opening up a little bit to show your followers you are a real person. People outside your field are not likely to follow you if your tweets are only about sharing events, articles, and positions in your own field. You need to add an extra ingredient—your opinions, or something personal—to what you share. One way to do this is through sharing failures: a rejected paper or job application, or even a spilt coffee. This is a great way to give and receive moral support from other academics.

Beyond academia, sharing something personal shows non-academics that scientists are just people too. You don't need to go into details of your life to do this—think about a nice book you have read recently, a concert you visited, or your cat, as long as you keep the vibe positive. Similar to [22], we think that humor and individuality should not impact the perceived value of science.

## Rule 10: Great power & great responsibility

Once you reach a substantial number of followers, let's say around 1,000, Twitter changes a bit from being a place where you go to share to a place where other people go to learn from you! This change can be a bit similar to going from PhD student to staff: suddenly you are (part of) defining what is and what isn't appropriate, interesting or cool.

This responsibility means that you will need to spend a bit more time to think about what you tweet. There are cases of people getting in trouble for things they have shared through Twitter—going as far as being taken to court for retweeting somebody else's opinion [23]. [8] also report an instance of having to seek help from the university's legal service. One challenge as you acquire more followers is to strike a balance between remaining open and personal, while making sure you do not upset too many people in power, whether inside or outside your own institution or field.

On the other hand, a large Twitter network also helps you to spread important ideas, and help people learn about opportunities you know about. To a certain extent, Twitter is an equalizer, whether you are a full professor or an undergraduate student.

You can also use Twitter to lift deserving people up. Ensure that you properly credit collaborators, mentors, and especially students to increase their visibility. A great hashtag to promote researchers is #ScholarSunday, an academic version of #FollowFriday created by Dr. Raul Pacheco-Vega.

## Conclusion

In this article, we have introduced the way in which researchers can benefit from using Twitter, and we listed a number of tips for effective tweeting, both for people who are considering joining Twitter and for those who are already ardent Twitter users.

As with any other aspect of human life, Twitter has trade-offs between quantity and quality. It is, unfortunately, a frequent practice to retweet any content spread by one's collaborators or friends without even reading it. This allows users with high numbers of followers to gain further attention with minimal effort in a kind of "snowball" effect. But chasing after numbers in this or other ways (e.g. tweeting with lots of hashtags in order to maximize your number of followers) is typically a short-sighted strategy, as Twitter is also inhabited by a large number of bots (automated accounts) which will tend to follow you and instead of contributing to your networking efforts, distract you by flooding you with a lot of unwanted content.

Twitter is also a public platform and one should be careful about spreading any content that is judgmental about other people, including their personality, their beliefs and their work. Some people have an inborn sense of diplomacy while others need to learn it the hard way. You might find it prudent to observe other researchers' interactions on Twitter for a while before you indulge in heated conversations on any topic.

Furthermore, one should remember that Twitter's business model is based on advertising. Therefore, the users need to be aware that some of the content is sponsored and aims solely at selling products to the user. As mentioned earlier, Twitter is also populated by bots which are designed to enable certain accounts to increase their numbers of followers. This often takes the form of a certain users following you and then unfollowing you immediately after you follow them back.

Regardless of these potential downsides, Twitter is still strongly recommended for anyone who needs to develop themselves in academia, to learn and to teach, to develop and tighten bonds with researchers overseas, and to join the #AcademicTwitter community.

## Acknowledgments

We would like to acknowledge all the inspiring people we have met on Twitter during our careers. We would also like to personally acknowledge Melanie Stefan (@MelanieIStefan) for encouraging us to write this piece, and for brainstorming with us and Stefanie Brackenhoff for assisting with the editing. Finally, we thank Nick Brown (@sTeamTraen) for correcting the English in a tweet from one of the authors and then volunteering to proofread more or less our entire manuscript.

## References

1. Bik H., Dove A., Goldstein M., Helm R., MacPherson R., Martini K., et al. (2015). Ten Simple Rules for Effective Online Outreach. PLOS Computational Biology, 11(4), e1003906. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439
2. Ekins S. & Perlstein E. (2014). Ten simple rules of live tweeting at scientific conferences. PLoS Computational Biology, 10(8): e1003789. https://doi.org/10.1371/journal.pcbi.1003789 PMID: 25144683
3. Duhigg C. (2013). *The Power of Habit*: *Why we do what we do and how to change.* London, United Kingdom: Random House.

4. Baker M. (2015). Social media: A network boost. *Nature* 518: 263–265. https://doi.org/10.1038/nj7538-263a PMID: 25679032
5. Lee J.-S. (2019). How to use Twitter to further your research career. *Nature*. https://doi.org/10.1038/d41586-019-00535-w
6. Lurie, D., & Kessler, D. (2018). On the nature of time-varying functional connectivity in resting fMRI. https://doi.org/10.17605/OSF.IO/XTZRE
7. Cheplygina, V. (2019). How I fail. Retrieved from http://veronikach.com/category/how-i-fail/
8. Britton B., Jackson C., & Wade J. (2019). The reward and risk of social media for academics. *Nature Reviews Chemistry*, 3:459–461. https://doi.org/10.1038/s41570-019-0121-3
9. Zhao, D., & Rosson, M. (2009). How and why people Twitter. Proceedings of the ACM 2009 International Conference on Supporting Group Work. https://doi.org/10.1145/1531674.1531710
10. Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter, 43rd Hawaii International Conference on System Sciences, Honolulu, HI, pp. 1–10. https://doi.org/10.1109/HICSS.2010.412
11. Daddow O. (2018) Why is Reviewer 2 always the one who says you should have written on a different topic using a different method? Asking for a friend [Twitter Post] https://twitter.com/oliver_daddow/status/1070305080711241729
12. Bourne P., & Barbour V. (2011). Ten Simple Rules for Building and Maintaining a Scientific Reputation. PLOS Computational Biology, 7(6), e1002108. https://doi.org/10.1371/journal.pcbi.1002108 PMID: 21738465
13. Crusoe, M. There is no scientific justification for harassment #bropenscience Anyone may block or mute anyone they wish without justification. No one is obligated to read any Tweet from any person. [Twitter Post] https://twitter.com/biocrusoe/status/1149297919528452097
14. Haynes, T. (2018, January 5). Dopamine, Smartphones & You: A battle for your time. [Blog post]. Retrieved from http://sitn.hms.harvard.edu/flash/2018/dopamine-smartphones-battle-time/
15. Price, C. (2018, August 8). The Real Cost of Phone Addiction. [Blog post]. Retrieved from Medium website: https://medium.com/s/greatescape/the-real-cost-of-phone-addiction-5a57a9b1be86
16. Horst M., Davies S., & Irwin A. (2016). Reframing science communication. In *Handbook of Science and Technology Studies* (pp. 881–908). Cambridge, United States: MIT Press.
17. Hermans, F. (2018). Programming by calculation. [Slides]. Retrieved from https://www.slideshare.net/Felienne/programming-by-calculation
18. Van der Zee, T. (2016). There are bad visualizations, and then there's the 'bicycle of education'. [Twitter Post] https://twitter.com/research_tim/status/737757291437527040
19. Albers, C. (2018). New preprint. Comments welcome. [Twitter Post] https://twitter.com/caal/status/966279936028958720
20. Priem J. and Costello K. (2010), How and why scholars cite on Twitter. Proceedings of the American Society for Information Science and Technology, 47: 1–4. https://doi.org/10.1002/meet.14504701201
21. Smeets I. (2014). Het exacte verhaal. Wetenschapscommunicatie voor bèta's (Science communication for scientists). Amsterdam, Netherlands: Nieuwezijds
22. Dashnow H., Lonsdale A., & Bourne P. (2014) Ten Simple Rules for Writing a PLOS Ten Simple Rules Article. PLoS Comput Biol 10(10): e1003858. https://doi.org/10.1371/journal.pcbi.1003858 PMID: 25340653
23. Fried, E. (2018) James Coyne sued me for cyberbullying. [Blog post]. Retrieved from https://eiko-fried.com/james-coyne-sued-me-for-cyberbullying/

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules to create biological network figures for communication

**G. Elisabeta Marai[1☯*], Bruno Pinaud[2], Katja Bühler[3], Alexander Lex[4], John H. Morris[5☯]**

**1** Electronic Visualization Laboratory, University of Illinois at Chicago, Chicago, Illinois, United States of America, **2** Laboratoire Bordelais de Recherche en Informatique, University of Bordeaux, Bordeaux, France, **3** Biomedical Image Informatics Department, VRVis Research Center, Vienna, Austria, **4** Department of Computer Science, University of Utah, Salt Lake City, Utah, United States of America, **5** Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, United States of America

☯ These authors contributed equally to this work.
* gmarai@uic.edu





**Citation:** Marai GE, Pinaud B, Bühler K, Lex A, Morris JH (2019) Ten simple rules to create biological network figures for communication. PLoS Comput Biol 15(9): e1007244. https://doi.org/10.1371/journal.pcbi.1007244

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** September 26, 2019



**Funding:** GEM's work was supported in part by US federal agencies, through awards NSF CNS-1625941, NIH NCI-R01CA214825, NIH NCI-R01CA2251, and NIH NLM-R01LM012527. KB's work has been funded by the Austrian Research Promotion Agency's FFG COMET - Competence Centers for Excellent Technologies program (project 854174). JHM's work has been funded by NIH P41-GM103504 and grant number 2018-183120 from the Chan Zuckerberg Initiative DAF, an advised fund of the Silicon Valley Community Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Biological network figures are ubiquitous in the biology and medical literature. On the one hand, a good network figure can quickly provide information about the nature and degree of interactions between items and enable inferences about the reason for those interactions. On the other hand, good network figures are difficult to create. In this paper, we outline 10 simple rules for creating biological network figures for communication, from choosing layouts, to applying color or other channels to show attributes, to the use of layering and separation. These rules are accompanied by illustrative examples. We also provide a concise set of references and additional resources for each rule.

## Author summary

Biological network figures are ubiquitous in the biology and medical literature. In this paper, we outline 10 simple rules for creating biological network figures for communication, from choosing layouts, to applying color or other channels to show attributes, to the use of layering and separation.

## Introduction

Biological networks are present in many areas of biology, including studies of cancer and other diseases, metagenomics, pathway analysis, proteomics, molecular interactions, cell–cell interactions, epidemiology, network rewiring due to perturbations or evolution, etc. Increasingly, published studies in these areas and many others include figures meant to convey the results of one or more experiments or of the network analysis carried out. As a result, biological network figures are ubiquitous in the biology and medical literature. On the one hand, a good network figure is able to quickly provide information about interactions between items and can often convey the nature and degree of interactions, as well as enable inferences about the reason for those interactions. On the other hand, good network figures are difficult to

**Competing interests:** The authors have declared that no competing interests exist.

create. The scale of data can often obscure the relationships that the figure is trying to convey, the spatial layout and distribution of the network can be difficult to interpret, and the many ways in which data can be mapped onto network representations provide an easy pathway to violating best practices of data visualization.

Some relatively simple rules, when followed, can significantly improve the likelihood that a network visualization will "tell the story" the author intends. The following set of rules was a result of a week-long seminar that brought together leading biology, bioinformatics, and visualization researchers from different countries [1]. Note that the rules we give are meant for static figures as used for publications, not for dynamic figures or for interactive or exploratory tools that allow users to manipulate the data view. The rules are tightly interconnected and, in general, follow the typical visualization design decision process (without forming a decision tree, due to their interconnectedness), from determining first the intended message of the illustration we seek to create [2, 3] to selecting appropriate encodings for that message and network. In order to provide a useful interpretation of these rules, we use real data for our illustrations below, and in many cases, we utilize network figures from the bioinformatics literature. In no way do we mean to detract from the science or experimental results that these published figures are trying to represent. As already noted, good network figures are difficult to create, and even some of the figures we use to illustrate specific rules below may come up short with respect to another rule. Last but not least, for each rule we also provide a concise set of references and resources where the interested reader may find additional information on the topic.

## Rule 1: First, determine the figure purpose and assess the network

The first rule is also arguably the most important: Before creating an illustration, we need to establish its purpose [4] and then the network characteristics. When establishing the purpose, it helps to first write down the explanation (caption) we wish to convey through the figure and note whether the explanation relates to the whole network; to a node subset in the network; to a temporal, causal, or functional aspect of the network; to the topology of the network; or to some other aspect. This analysis needs to happen before we draw the network because the data included in the view, the focus of the figure, and the sequence we use to visually encode the network should support the explanation that we wish to convey. For example, salient aspects of the figure may need to be displayed centrally, in larger size, or marked by annotations. Second, we need to assess the network in terms of scale, data type, structure, etc. These network characteristics will further constrain salient aspects of the visualization, such as the color, the shape, the marks used, and the layout of the network [5].

Fig 1 delivers two messages about proteins known to be involved in glioblastoma multiforme (GBM). The first figure is a RAS signaling cascade in a curated GBM network. Because the message of the figure relates to protein interaction functions, the figure uses a data flow encoding, with nodes connected by arrows. The nodes are colored by the expression variance across samples. The second figure is a STRING protein–protein interaction (PPI) network representing proteins that show significant expression changes in subtype 3 of GBM, in addition to 20 additional proteins to improve connectivity; the colors represent the fold change, and the size represents the number of mutations. Because the message of this figure relates to the structure of the network, not its functionality, the nodes are connected by undirected edges, and the nodes are placed to reinforce the structure. Furthermore, note how the quantitative color scheme (yellow to green gradations) in the first network shows expression variance, whereas the divergent color scheme (red to blue) in the second network emphasizes the extreme values of differential expression for one GBM subtype. Similarly, the edges in the first network are arrows indicating function, whereas in the second network, they are edges to indicate

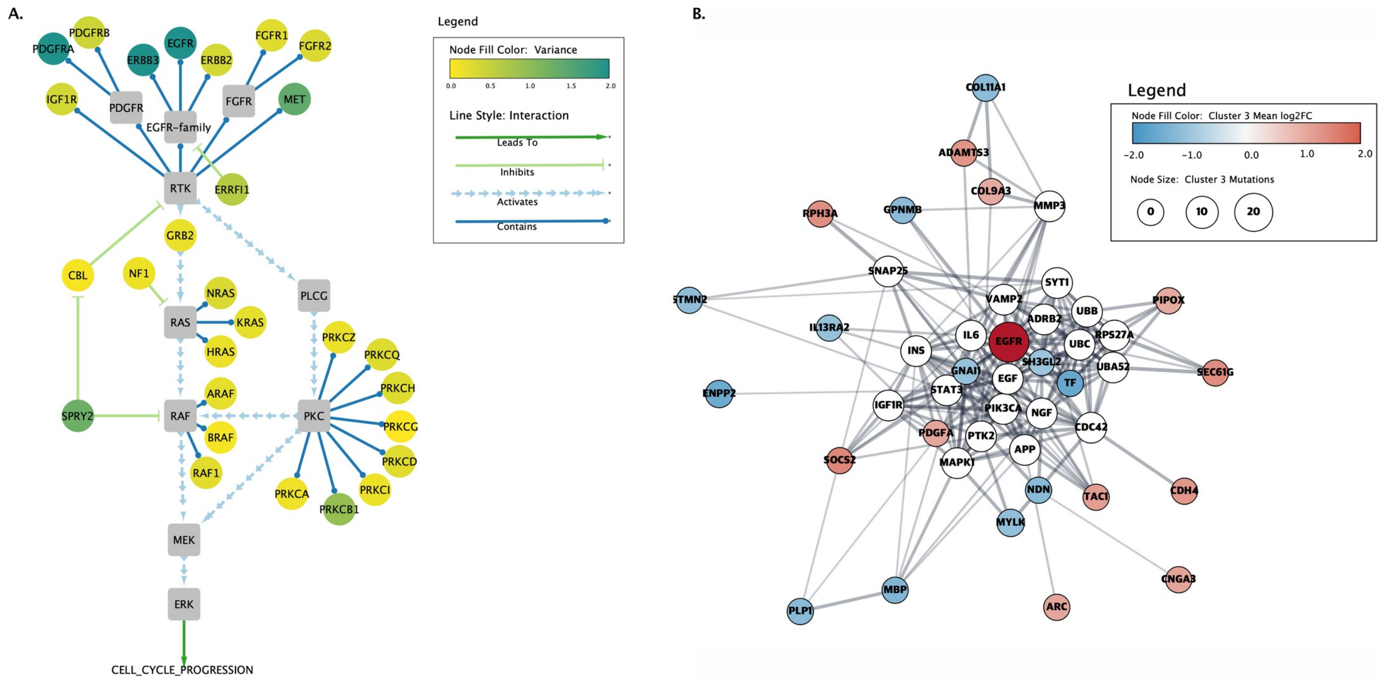



**Fig 1. First, determine the figure purpose and assess the network.** Two representations of proteins involved in GBM. The left image (A) shows a curated cancer signaling pathway taken from the TCGA's original Mondrian plugin to Cytoscape (Cytoscape Consortium; https://cytoscape.org/). The node color represents the overall variance of expression across a set of patients, and the lines and arrows represent the function of the interactions between the proteins. In the right image (B), a PPI network was created using the Cytoscape stringApp and annotated with data downloaded from TCGA. The colors represent the fold change for subtype 3 of GBM, the node sizes vary with the number of mutations, and the edges represent functional associations. GBM, glioblastoma multiforme; PPI, protein–protein interaction.

https://doi.org/10.1371/journal.pcbi.1007244.g001

structure. Each image tells a different story: The first message is about network functionality, the second about the network structure.

## Rule 2: Consider alternative layouts

Node-link diagrams are the most common way to display network data. Node-link diagrams are familiar to readers, and they can show relationships between nodes that are not immediate neighbors. However, node-link diagrams also have drawbacks: For dense and large networks, they tend to produce significant clutter, edge attributes are difficult to visualize, and node labels often cause even more clutter. An alternative network representation is adjacency matrices (see Fig 2). An adjacency matrix lists all nodes of a network horizontally and vertically. An edge is represented by a filled cell at the intersection of the connected nodes.

Adjacency matrices have several advantages: First, they are well suited for dense networks with many edges, as every possible edge is represented by a cell [7]. Second, they can encode edge attributes, for example, with color or color saturation of a cell. Third, adjacency matrices excel at showing neighborhoods of nodes and clusters, provided the node order is optimized [8]. Fourth, the layout of the matrix makes it easy to display readable node labels, whereas labels in a comparable node-link layout would cause significant clutter. Matrix layouts are easy to implement, e.g., in R, Python, or JavaScript, even without dedicated graph visualization libraries. In practice, using an appropriate column/row reordering algorithm is crucial [8].

Another alternative to traditional node-link layouts is fixed layouts: Here, the nodes are positioned such that the position of the nodes themselves encodes data. A common example is

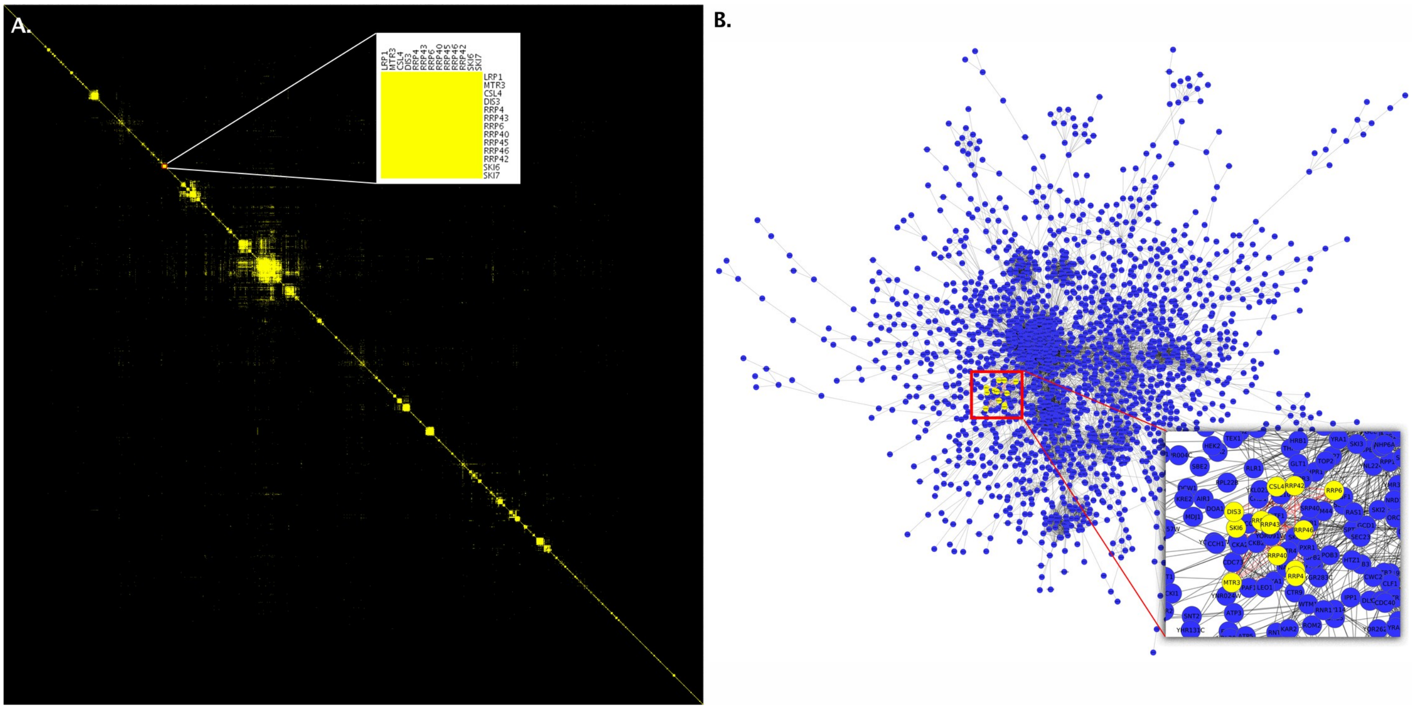



**Fig 2. Consider alternative layouts.** These two images represent the same data from Collins and colleagues [6]. The image on the left (A) shows an adjacency matrix representation of the network. The inset within the image shows a cluster identified on the diagonal that represents the exosome complex. The image on the right (B) is of the same data depicted as a node-link diagram with the same nodes highlighted. Notice how difficult it is to see the close interaction between the nodes, even in the inset in this second image, due to the clutter resulting from other nodes. These images were produced in Cytoscape (Cytoscape Consortium; https://cytoscape.org/) with the clusterMaker2 app and postprocessed in Photoshop (Adobe; https://wwww.adobe.com/) to merge in the insets.

https://doi.org/10.1371/journal.pcbi.1007244.g002

networks shown on top of maps, or links on top of linear or circular layouts, such as is commonly used for genomic data visualization in Circos [9]. Finally, when the graph to be shown is a tree, we can also make use of implicit layouts, such as icicle plots [10], sunburst plots [11, 12], or treemaps [13, 14]. Implicit layouts encode the relationships between parents and children by adjacency, and the size of the leaves is commonly scaled according to an attribute. S. Ribecca's Data Visualisation Catalogue (datavizcatalogue.com) provides a wide although non-exhaustive array of possible representations.

## Rule 3: Beware of unintended spatial interpretations

Node-link diagrams map nodes to locations in space. In turn, Gestalt theory (in particular, the principles of grouping) teaches us that the spatial arrangement of nodes and edges influences the reader's perception of the network information—even if there is no meaning [4]. Thus, the right layout can effectively enhance features and relations of interest, but the wrong layout might easily lead to misinterpretation. An example of such a misinterpretation can be found in the Atlas of Science [16]. Although aesthetically pleasing, the node-link diagram shows a defective spatial encoding that suggests a black hole of knowledge.

Proximity, centrality, and direction of node arrangement are the most prominent principles to be considered when integrating spatiality into meaningful network representations: Nodes drawn in proximity will be interpreted as conceptually related; nodes grouped together are also perceived as more similar to each other than nodes outside the group. We may use as a similarity measure the connectivity strength between two nodes (an edge-based measure), similarity of the content carried by the nodes, e.g., nodes being part of the same brain region or conceptual group (a node-based measure), or a mixture of both. This measure is then used as an optimization criterion for the layout algorithm (Fig 3). Most prominent layouts are force directed and interpret the given similarity measure as an attracting force for nodes, whereas graph layouts based on multidimensional scaling perform better for cluster detection [15]. Centrality is a design principle in which the center and periphery may represent metaphorically high relevance and secondary relevance, respectively. A layout may be spatially constrained to display the focus of the illustration in the center of the figure. The third design principle is direction: The vertical dimension represents power, from light/good (up) to heavy/bad (down) and also flow of information or development (up to down) or in the horizontal direction (left to right in Western cultures).

Most open-source network drawing tools like Cytoscape (Cytoscape Consortium; https://cytoscape.org/) and yEd (yWorks GmbH; https://www.yworks.com) provide a rich selection of different layout algorithms. Beside these resources, drawing networks and developing appropriate layout methods is a whole scientific discipline by itself. An excellent source for diving deeper into the world of graph drawing algorithms is http://graphdrawing.org/.

## Rule 4: Provide readable labels and captions

The proper use of labels and captions can help explain and clarify the icons, colors, and visual representations present in a network figure. First, network labels and, in general, text in a network figure have to be legible. To be legible, labels in the figure should use the same (or larger) font size as the caption font, not smaller. Fig 4A shows PPI data from Andrei and colleagues [34], in which the node labels are too small to be legible. In Fig 4B, the layout has been modified to make better use of the available space, resulting in larger labels. Although this type of manipulation may not always be possible (for example, Fig 10 in Wenskovitch and colleagues [19] shows the similarity among 4 large-scale network models with no room for larger labels), in such cases, one should at least provide an online high-resolution version of the network that

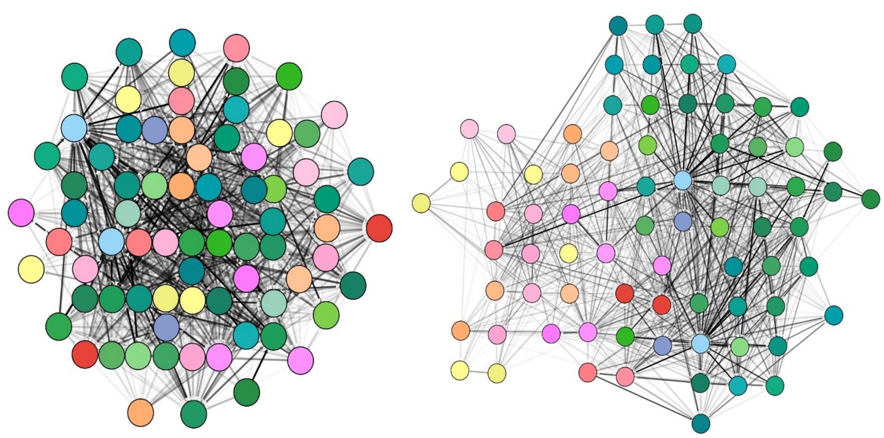


**Fig 3. Beware of unintended spatial interpretations.** This figure shows two illustrations representing the same region of the normalized structural mouse brain connectivity data set described by Ganglberger and colleagues [17]. The data are derived from the Allen Mouse Brain Connectome dataset [18]. The illustrations have been generated using the Cytoscape.js (Cytoscape Consortium; https://cytoscape.org/) implementation of the force-directed layout algorithm CoSE. The left image uses connectivity strength as the driving force for the layout, posing strongly connected nodes closely together, but at the same time neglecting the spatial context of the network. Instead, the second layout in the right image is driven by the spatial relation of brain regions, generating automatically a "flattened" mouse brain representation as seen from above. Symmetry and spatial positions are approximately reproduced. Structural connectivity strength is encoded by the gray-level color scale of the edges. CoSE, compound spring embedder.

https://doi.org/10.1371/journal.pcbi.1007244.g003

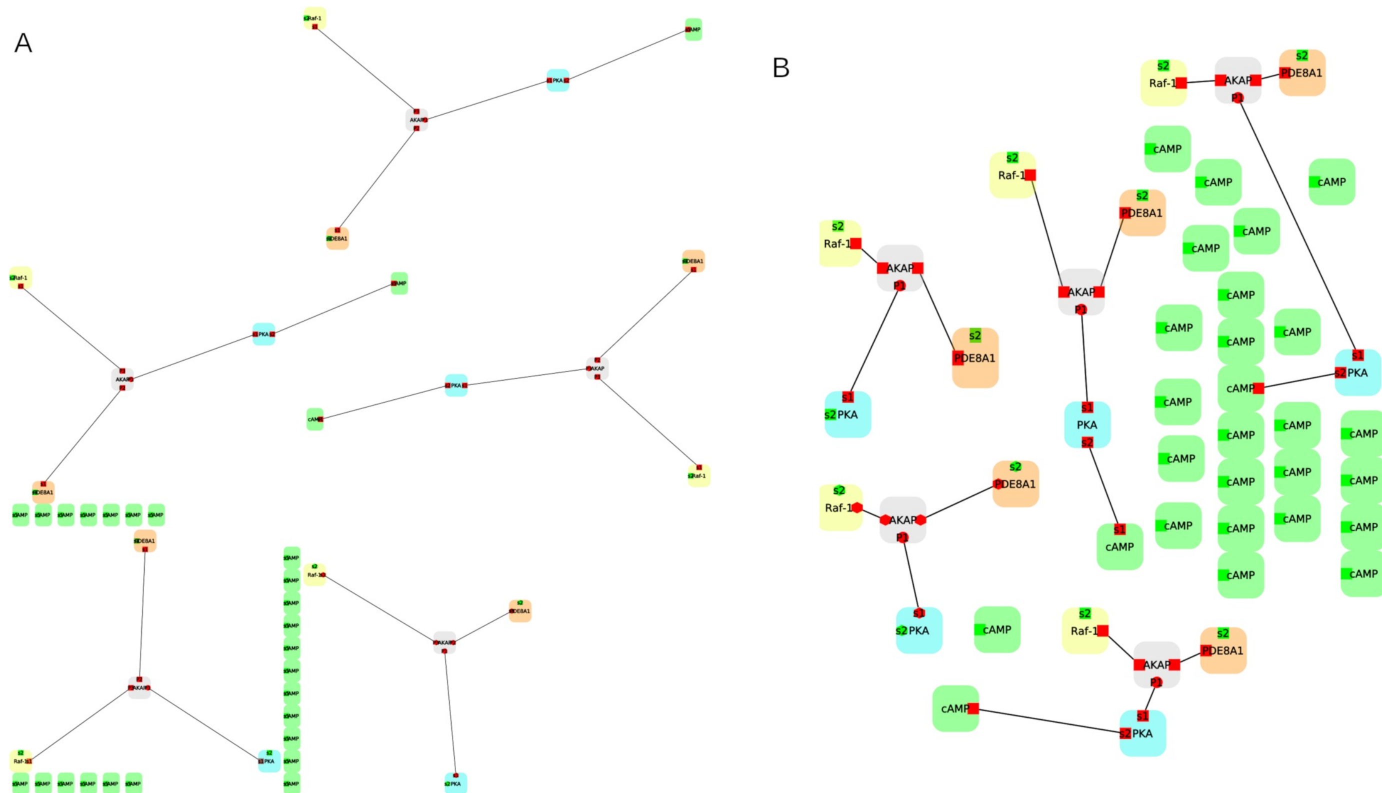



**Fig 4. Provide readable labels and captions.** (A) An example network based on PPI data from Andrei and colleagues [34], in which the node labels are too small to be legible. (B) The same network, but this time the layout has been improved to make better use of the available space, resulting in larger labels. The two images have been generated using the open-source software Porgy (http://porgy.labri.fr). PPI, protein–protein interaction.

https://doi.org/10.1371/journal.pcbi.1007244.g004

can be zoomed in. Furthermore, whereas it is tempting to rotate text affiliated with specific network elements in order to optimize space, all network text should use a horizontal orientation: Vertical or tilted text is hard to read. To be legible, all text should also have good contrast with the background, preferably black on white or white on black.

The figure and its caption (the brief explanation appended to an image) should each be able to stand on their own and provide both context and interpretation. The caption, in particular, should tell the reader what to notice in the network figure, without the reader needing to chase the figure reference in the manuscript text. The network figure text should further clarify the meaning of all unusual visual markers and channels used in the network representation, including all colormaps. Last but not least, labels should be properly placed within the network figure. For example, inset and subfigure labels should be placed in clear proximity to that element. Whenever possible (i.e., when the figure is not too cluttered), use direct labeling instead of numerical pointers to a legend; numerical pointers place a higher cognitive load on the reader.

## Rule 5: Choose the right level of detail

Depending on the intended meaning of a figure, it may be beneficial to show fewer details, even if they are relevant, in order to bring into better focus the item(s) or relationship(s) of interest (reference [5], Chapter 13). The level of detail shown can also change locally across the figure. If, for example, one is interested in showing centrally the details of a network, there is no need to display the data at the periphery with the same (high) level of detail. To keep the context of the visualization clear, the entire structure can be shown in an aggregated form, around the item of interest. Aggregation can be performed at the level of items, based on dimensionality reduction over the item attributes (e.g., principal component analysis), or based, for example, on a spatial aggregation of geo-collocated items into groups. Aggregation may also be performed at the level of relationships, via, e.g., edge bundling algorithms. The wise use of aggregation in combination with a variety of visual marks and channels can significantly reduce visual clutter.

Fig 5 shows images made with Cytoscape of protein interaction data with 5 complexes (computationally determined) colored using data from Kuhner and colleagues [21]. This figure replicates the sequence of steps described in Gehlenborg and colleagues [20]. Network *a* is the original protein interaction network (> 400 proteins). According to Gehlenborg and colleagues, this first network is hardly readable, and nothing really interesting is visible. Network *b* is a recomputed network after removing nodes not of interest. Clusters based on the complexes' color start to emerge. Network *c* is a manual refinement to emphasize the structure of protein complexes and the interactions between them. Finally, network *d* proposes to collapse nodes in each complex core (e.g., nodes inside each colored circle are replaced by only one triangle of the same color) to simplify the network and emphasize global properties, which is the aim of the figure.

## Rule 6: Use color responsibly

Color is a complex topic [23], and here, we touch only on the aspects most relevant to bionetwork visualization. Color is a perception and not visible electromagnetic radiation (light waves are not colored): Most, but not all, people experience the sensation "blue" with wavelengths near 400 nm. The color humans perceive depends on the eye–brain mechanism, and therefore color perception is influenced by context, training, or abnormalities such as color blindness, which affects 8% of the male population and often results in an inability to distinguish red from green. For this reason, red–green color encodings of network data should be avoided. Human vision is also much more sensitive to slight changes in the luminance of a color (its intensity or value) than slight changes in the quality of a color (its hue and saturation) [24]. Therefore, it is a good idea to convert the network figure to grayscale and make sure that the information encoded in the diagram is still legible. In a nutshell, get the figure right in grayscale first. In terms of saturation, areas of saturated color draw attention and are best used on small areas such as nodes; use saturated colors sparingly and to draw attention. The hue component (the color quality that distinguishes red, green, blue, etc.) is also powerful: Hue families can code related items. Qualitative maps (i.e., multihued maps) should be used only for categorical coding, to indicate different qualities or identities of data. Because humans have no sense of whether blue is more or less than orange, to encode ordinal data, figures should use a progression of luminance values, similar to topographic maps. All else being equal, blue-family hues tend to recede, whereas warmer red-family hues tend to come forward, and so the use of these two families together in a network may result in an unwanted 3D effect [25]. Transparency can be further used to modulate a color: Transparent markers tend to be perceived as being in the background.

In Fig 6A, the colormap encodes the node degree using a two-tailed gradient (saturated yellow to saturated green) and saturated red for 1. The color scheme is not color-blind safe and employs saturation incorrectly. Some edges use, confusingly, the same hue as some unconnected nodes. The gray figure text has also poor contrast with the background (i.e., the text and background have similar luminance), making it hard to read. The revised image in Fig 6B uses a ColorBrewer (http://www.colorbrewer2.org) sequential colormap for the node degree, a

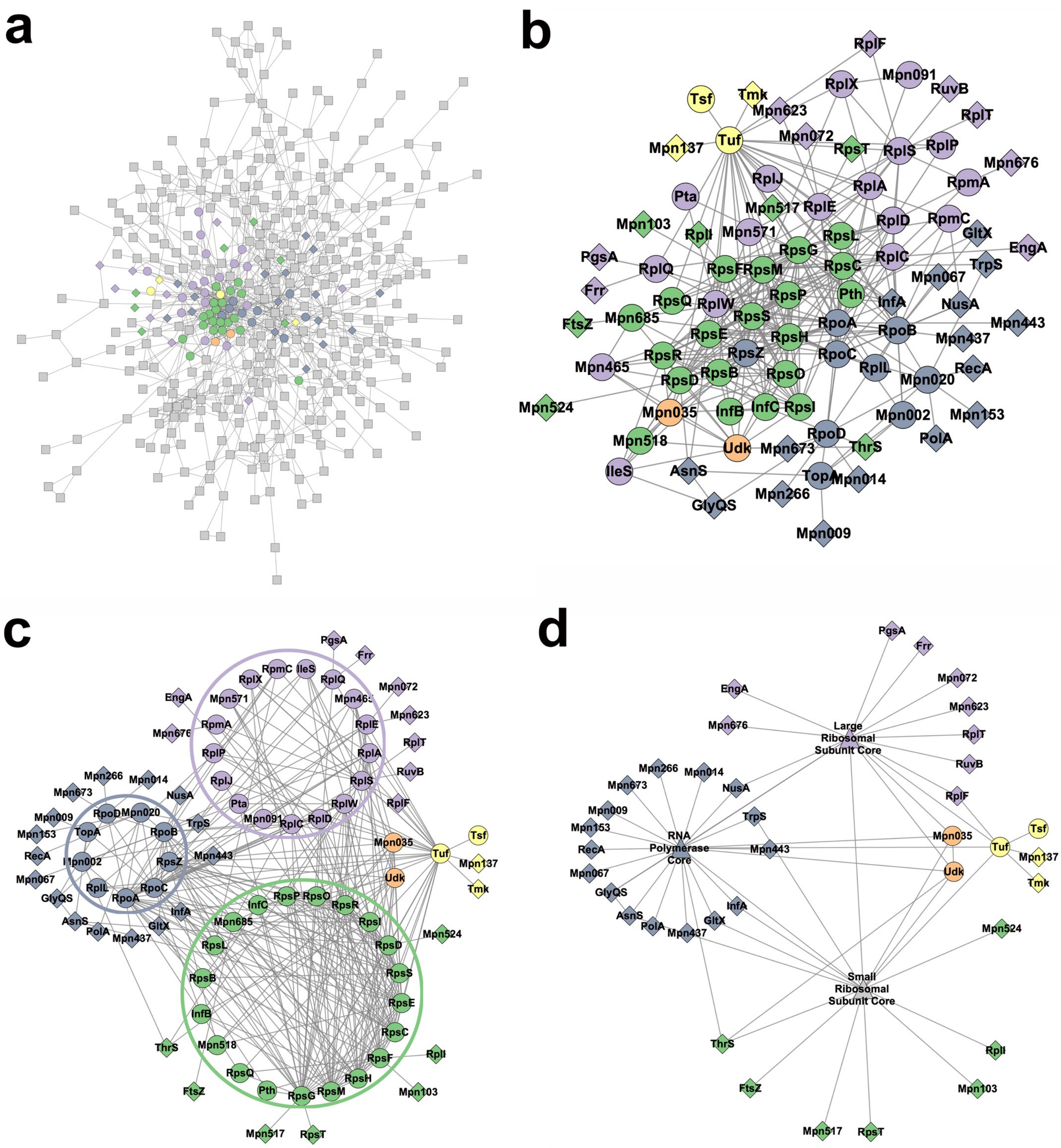



**Fig 5. Choose the right level of detail.** Example aggregation using data from Kuhner and colleagues [21], which replicates the sequence of steps described in Gehlenborg and colleagues [20], from a hardly readable network (A), gradually through (B) and (C), to a legible, aggregated version of the same network (D).

https://doi.org/10.1371/journal.pcbi.1007244.g005

separate sequential colormap for edges, and black figure text. The result is a significantly clearer figure, although the text contrast with colored backgrounds could be further improved.

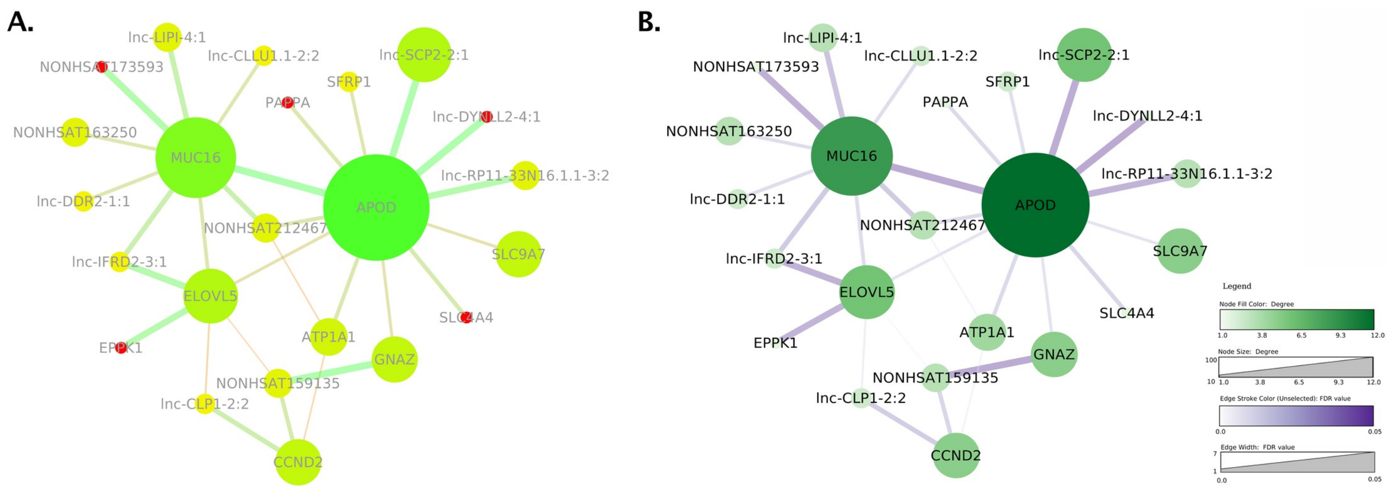



**Fig 6. Use color responsibly.** Two network images based on data from Khaled and colleagues [22]. (A) is a recreation of the original Fig 3B shown in the paper, including the color-blind and saturated color scheme, which makes it difficult to perceive the relative importance of the nodes. The colormap also groups unrelated edges and nodes together through similar colors, whereas the node labels in light gray have low luminance contrast with the white background and are difficult to read. (B) shows an improved version, including a legend and appropriate and separate quantitative colormaps for edges and nodes. Both images were created with Cytoscape (Cytoscape Consortium; https://cytoscape.org/) and postprocessed using Photoshop (Adobe; https://www.adobe.com/) to assemble them.

https://doi.org/10.1371/journal.pcbi.1007244.g006

## Rule 7: Use other visual marks and channels appropriately

Whereas color is incredibly powerful, other visual marks and channels are also important. Marks are basic geometric elements that depict items or links, whereas channels control the appearance of marks. Marks can be, with increasing dimensionality, dots, lines, arrows, blobs or polygons (marks with area) or volumetric glyphs (marks with volume). Some channels are position (see Rule 4), color (see Rule 6), shape, size, tilt, area, and volume. Using a variety of marks wisely can create more powerful displays, through increased flexibility, and further allows layering and separation of information for more effective displays (Rule 8). With respect to marks, in general, dots and glyphs represent items, whereas lines and arrows represent relationships between items. Blobs represent regions or containers of items. Arrows are asymmetric lines that represent asymmetric relations and can change drastically the meaning of a figure: diagrams with arrows tend to be interpreted as functional, presenting a sequence of actions and outcomes. In contrast, diagrams without arrows tend to be interpreted as structural, specifying the location of parts relative to one another [4]. With respect to channels, position, color, and shape are identity channels, which means that a set of shapes can be used to distinguish different categories and so can a set of colors or a set of predefined positions [5]. The remaining channels are magnitude or quantitative channels, which means that a set of sizes (small, medium, large, etc., or weak, medium, strong, etc.) can be used to distinguish different quantities or attribute strength of a specific category, and so on. The example in Fig 7 shows network data from Morris and colleagues [26] and makes effective use of multiple visual marks and channels.

## Rule 8: Use layering and separation

The goal of any figure is to communicate information. Communication can be difficult if the key information is obscured by too much "clutter." We can raise the prominence of key information by imagining that different classes of information belong in different layers and that the key information is sitting on a higher layer in the figure and by providing visual separation

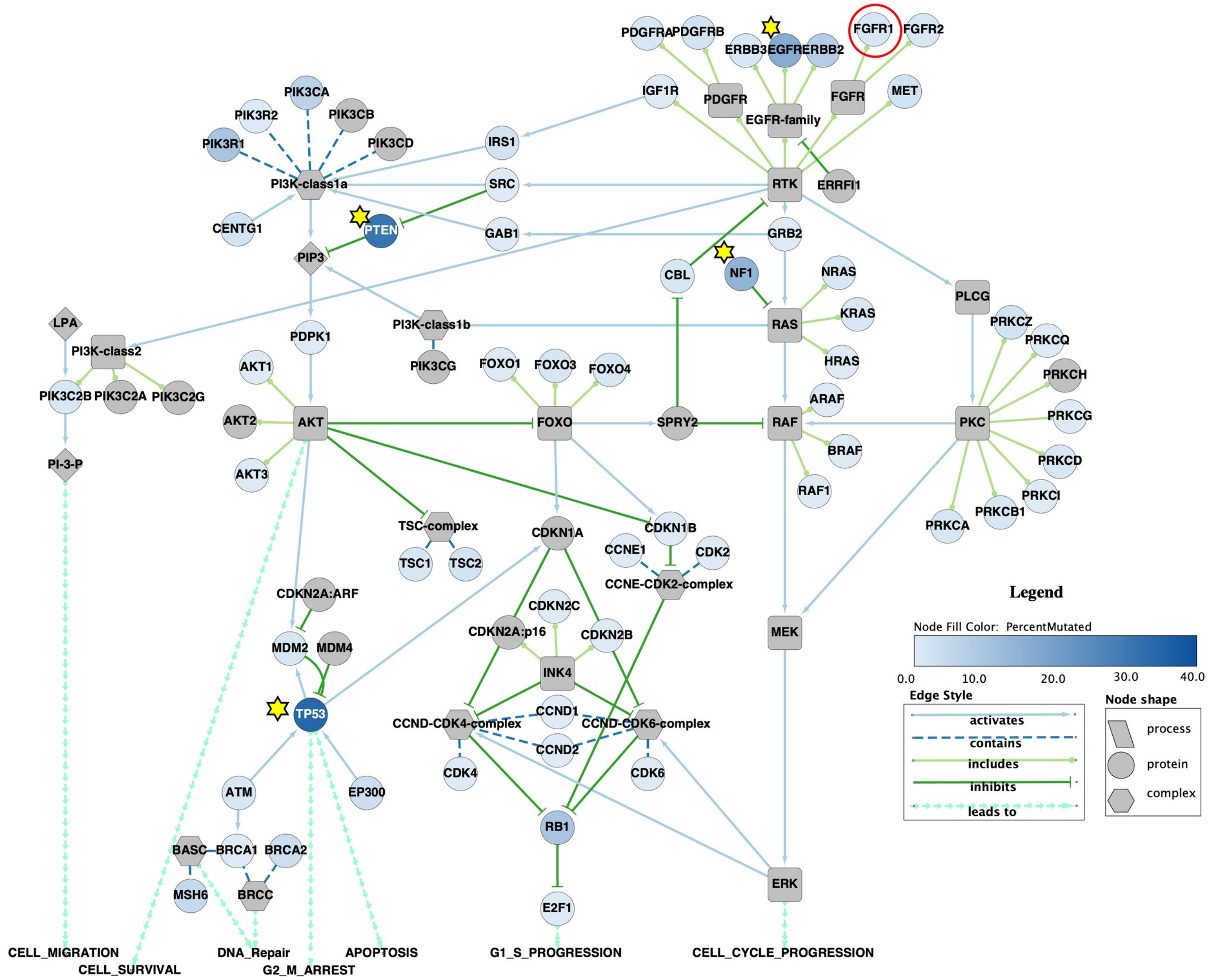



**Fig 7. Use other visual marks and channels appropriately.** In this Cytoscape (Cytoscape Consortium; https://cytoscape.org/) recreation of Fig 3 from Morris and colleagues [26], the authors used several different marks to explain the data in the network, including stars to indicate highly mutated nodes (in addition to the color gradient) and a red circle to indicate the subject of one of the scenarios outlined in the paper. The authors also used different node shapes to distinguish among complexes, proteins, and processes, and different line and line ending styles to indicate the relationship among the nodes.

https://doi.org/10.1371/journal.pcbi.1007244.g007

between the layers. Once we decide on how we would like the information organized, layering and separation [27] are traditionally accomplished by means of assigning a specific weight, color, opacity, or size to each layer of information although we can also use spatial cues such as grouping to highlight relationships. For example, we can decrease the weight, luminance, saturation, opacity, or size of less important information, and increase the weight, luminance, saturation, opacity, or size of the key information to make it more visually salient.

As an example, consider the images in Fig 8. The left image is a reconstruction of Fig 5A from Preston and colleagues [28], showing the largest subnetwork resulting from a pathway and enrichment analysis. Based on the callouts, the key data the authors want to convey are the neighborhoods around SRSF2 and NTRK1. The image on the right is an improved version in which we decreased the weight of those edges that do not connect to the key nodes and

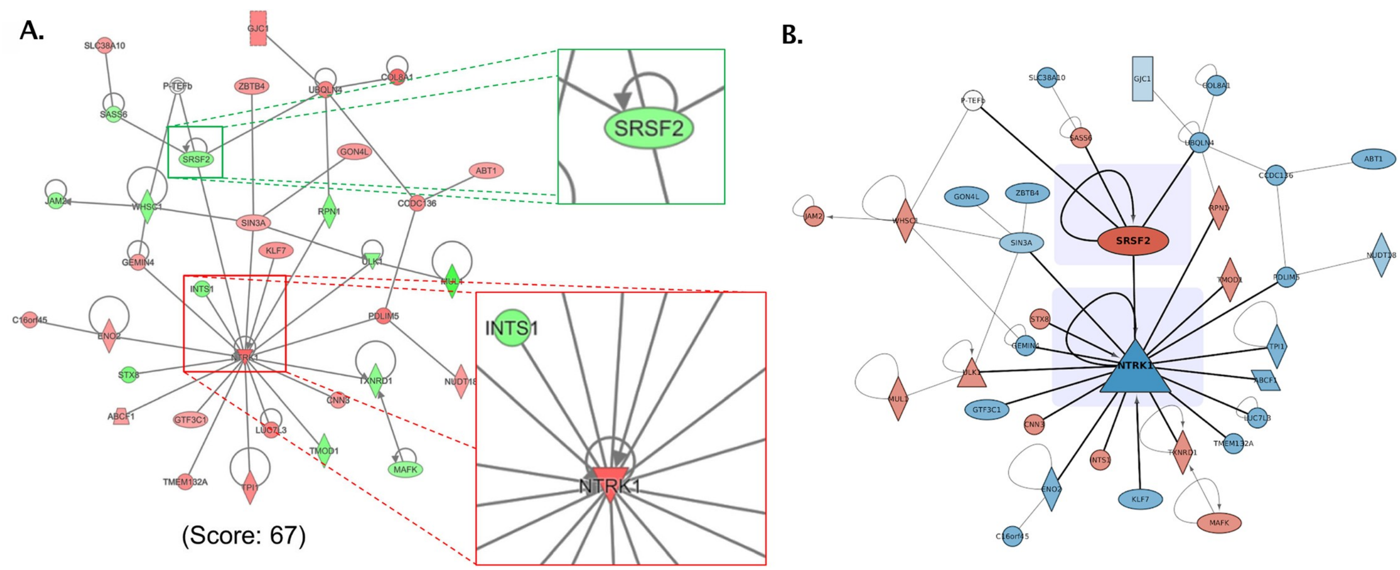



**Fig 8. Use layering and separation.** (A) Reconstruction of Fig 5A from Preston and colleagues [28], which contains the largest subnetwork resulting from a pathway and enrichment analysis. Callouts call attention to the neighborhoods around SRSF2 and NTRK1. (B) Modified image after changing the color scheme to avoid color-blind issues, decreasing the weight of the edges that do not connect to the key nodes and increasing the size of the key nodes. Nonkey nodes and self-edges were also de-emphasized by making them slightly transparent. Subtle shading behind the two key nodes was applied to provide additional separation.

https://doi.org/10.1371/journal.pcbi.1007244.g008

increased the size of key nodes (Rule 7). Nonkey nodes and self-edges were also rendered transparent, which effectively leads to a perception of these nodes and edges being in the background (Rule 6). Typically, if self-edges are not germane to the point being made by the image, they would be removed. Last but not least, subtle shading behind the two key nodes was applied to provide additional separation.

## Rule 9: Use multiple figures

Another kind of clutter in a network figure happens when there is too much information vying for the attention of the viewer. Under these circumstances, it is often better to split that information into multiple figures, each emphasizing a different point. Multiple figures can also effectively illustrate a sequence in the illustration. Thus, as a rule of thumb, count the number of visual properties an image uses to map data. If it is greater than 3, and they are not redundant (i.e., not intentionally mapping the same value for emphasis) and their interaction is not the point being made (i.e., overexpressed genes are also hubs), think about separating the image into multiple separate figures, each one emphasizing a different point and potentially focusing on relevant subnetworks. Another interesting aspect is the use of one image (e.g., A in Fig 9B) to provide overall context for the visualization of subnetworks. This overview + detail approach can be very useful. However, an extremely dense network with many overlapping nodes will not provide effective overview or context. Alternative models to the "overview-first" paradigm [30] include a "search-first" paradigm [31] and a "details-first" paradigm [32], depending on the interests and background of the target audience.

As an example, Fig 9A shows an image constructed from the data provided by Zhu and colleagues [29]. The "overview" network (A) is itself a 51-node subnetwork of the full 195-node network that the authors initially queried. This image includes several different pieces of information: The node colors indicate whether the node is a hub, square nodes represent a cluster found by the molecular complex detection algorithm (MCODE), and the purple borders

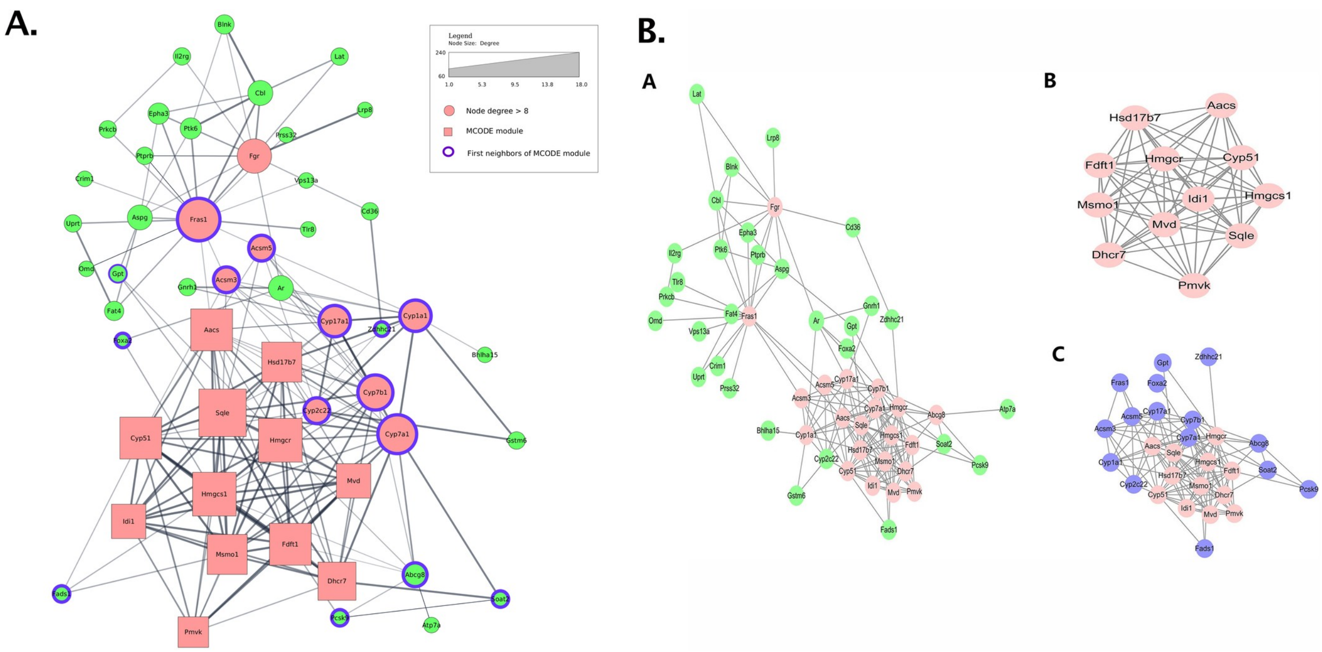



**Fig 9. Use multiple figures.** (A) An image constructed from the data provided by Zhu and colleagues [29] but constrained to show everything in a single view. The result is a very confusing image, and, from the viewer's perspective, it is hard to determine what is important. (B) The original image from Zhu and colleagues. Fig 5, where the authors split the network into 3 views, each view with a different focus. The first view (A) highlights the high degree nodes, the second view (B) shows the MCODE component, and the third view (C) adds the first neighbors to that component. MCODE, molecular complex detection algorithm.

https://doi.org/10.1371/journal.pcbi.1007244.g009

indicate the first neighbors of that cluster. The result is a confusing image, in which it is hard to determine what is important—the information does not rise above the clutter. Now, consider Fig 9B, which was the image the authors used. They split the network into three views. The first figure uses color to show degree, and it also provides an overall context for the subnetworks. The second network shows the results of the MCODE algorithm, and the third network shows those nodes plus their first neighbors. In each case, it is much easier to determine the point of the image.

## Rule 10: Do not use unjustified 3D

Many people think that if two dimensions (2D) are good, three dimensions (3D) must be better. As the printed medium evolves, video recordings and interactive displays, including virtual reality technologies, also become of interest. However, in the context of biological network displays, it is important to be aware that depth has important differences from the other two planar dimensions. 3D is seldom appropriate for such displays, due to documented issues related to depth perception inaccuracies, occlusion, perspective distortion, and so on (reference [5], Chapter 3). 3D is easy to justify when the users' tasks involve 3D shape understanding, for example, in molecular structures, which inherently have spatial structures. In such cases, the benefits of 3D absolutely outweigh the perception costs, and designers are justified in investing in interaction idioms designed to mitigate such costs. For example, occlusion hides information—some objects cannot be visible because they are hidden behind other objects. Even though the occluded nodes can be discovered via interactive navigation, the navigation has a time and cognitive cost. Occlusion can be also mitigated through the use of motion parallax (motion cues) [33], which also has an associated cost. In all other contexts, using 3D needs to be carefully justified in the context of the higher cognitive costs. As shown in the previous rules, there are other, more convenient techniques available for handling large scales, for example, avoiding showing an overview of the entire network altogether or choosing an alternative representation (e.g., an adjacency matrix) instead of node-link diagrams.

The example in Fig 10 shows a network illustration in which the height of each 3D cylinder is mapped to the size of specific network attributes. Note how the different cylinder heights can be mistakenly perceived as perspective foreshortening instead of different attribute sizes. A clearer illustration would use 2D instead and map the attribute size to a visual channel like 2D marker size.

## Conclusion

Several of the examples shown in this paper illustrate the many inherent difficulties in creating biological network figures that are appropriate for communication. The 10 simple rules we outlined in this paper show ways to improve such figures and in several cases, also illustrate the variety of means to visually encode information that circumvent data constraints. We believe these rules will benefit researchers who handle biological networks, be they bioinformaticians, neuroscientists, clinicians, and so on.

We strongly believe that creation of a biological network figure should start with an analysis of the intended figure message (Rule 1). Ideally, this analysis should be performed in conjunction with the domain scientists who generated the network data and its interpretation. Choosing an appropriate basic representation (node-link, matrix, etc.) and layout of the data comes next (Rule 2 and Rule 3), along with the appropriate labels and clarifying text (Rule 4). Gradual data preprocessing through aggregation (Rule 5), appropriate color mappings (Rule 6), the use

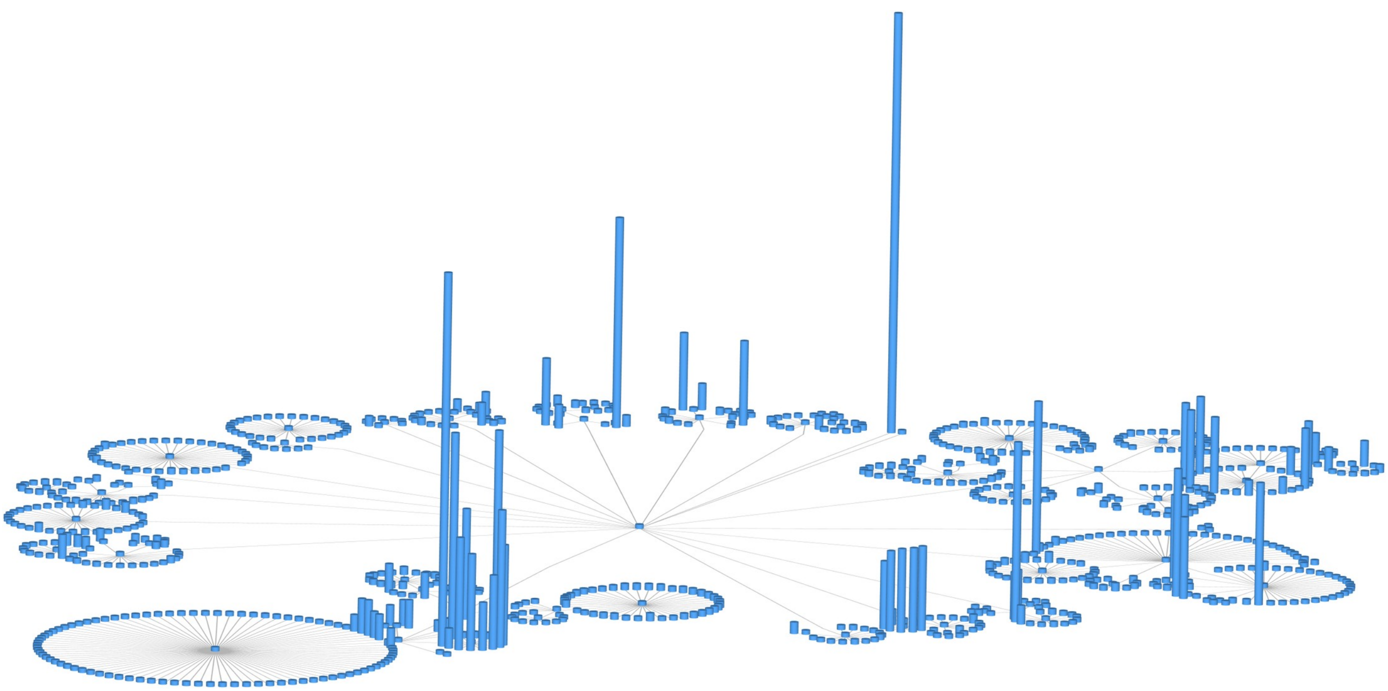


**Fig 10. Do not use unjustified 3D.** A 2D network displayed along an additional dimension in 3D. The height of each 3D cylinder is mapped to the size of a network attribute. Note the significant number of occlusions. This figure was generated using the open-source software Tulip (see the online Tulip user documentation, Chapter "Tulip in Practice: Four case studies" http://tulip.labri.fr).

https://doi.org/10.1371/journal.pcbi.1007244.g010

of an appropriate variety of marks and channels (Rule 7), layering and separation (Rule 8), and sequencing information along several figures (Rule 9) can then help reduce visual clutter and effectively emphasize the message of the figure. With advancements in media technology, we believe 3D figures should be used extremely cautiously, due to documented issues in depth perception (Rule 10).

An important aspect of network visualization that we have shown implicitly, although not discussed directly, is the power of network images to support the integration of a wide variety of data and to encode that data in a number of ways (for example, mapping expression fold change onto node fill color). This is an important and powerful feature of network visualization, particularly for exploring the results of multiple experiments in a single visualization in order to find new hypotheses or to confirm hypotheses, as often done in environments such as Cytoscape. On the other hand, too much information mapped onto a single figure can obscure the key aspects of that figure (see Rule 9), so it is important to balance how much of the network image is about the topology of the network and how much is about the integration of other -omics results in the context of gene or protein relationships. Fittingly, this observation rounds back the discussion to Rule 1—we first need to determine the purpose of the figure.

Another important aspect of network visualization that we have implicitly discussed is the issue of subnetworks. Whereas our rules suggest providing less detail at the periphery of a network, a periphery subnetwork may still be of major interest. Such situations may be addressed through the careful application of Rule 1 (determine first the message of the figure), Rule 8 (use layering and separation) to emphasize the subnetwork, and if necessary, Rule 9 (use multiple figures) to allocate a separate figure to that subnetwork.

Many of the illustrations in this manuscript have been generated using the open-source software platform Cytoscape. Wherever possible, we provided references to the software packages, as well as specific instructions. In an effort to make the application of these rules more accessible, we also provide, wherever possible, the session files for generating these images in a public repository (http://github.com/uic-evl/10RulesBionets). Whereas obviously there are many other software tools for network visualizations, we hope that knowing how to implement these rules in one tool might help the reader more easily transfer that knowledge to another tool. Beyond the basic "how-to" mechanics of the rules, we further that recommend biology researchers contact the biological data visualization community (e.g., http://biovis.net, http://bivi.co, http://visguides.org) for expert advice and help.

We trust that this minimal set of rules helps demystify the process of creating quality static biological network illustrations for communication. Although the landscape of visualization design is far more complex than briefly discussed in this paper, we hope this discussion clarifies some of the most common issues that arise in the creation of network figures, along with basic guidelines to help address those issues. We hope the interested reader will pursue the additional resources and references we include under each rule.

## Acknowledgments

We thank Dagstuhl Seminar 18161 (http://www.dagstuhl.de/18161) for the original inspiration for this work.

## References

1. Aerts J, Gehlenborg N, Marai GE, Nieselt KK. Visualization of Biological Data—Crossroads (Dagstuhl Seminar 18161). Dagstuhl Reports. 2018; 8(4):32–71. https://doi.org/10.4230/DagRep.8.4.32
2. Rougier N.P.,Droettboom M., Bourne P.E. Ten simple rules for better figures. Public Library of Science 2014. https://doi.org/10.1371/journal.pcbi.1003833

3. Lortie C.J. Ten simple rules for short and swift presentations. Public Library of Science 2017. https://doi.org/10.1371/journal.pcbi.1005373
4. Nathalie Henry Riche; Christophe Hurter; Nicholas Diakopoulos; Sheelagh Carpendale. Data-Driven Storytelling. AK Peters; 2018.
5. Munzner T. Visualization Analysis & Design. A K Peters Visualization. CRC Press; 2014.
6. Collins SR, Kemmeren P, Zhao XC, Greenblatt JF, Spencer F, Holstege FCP, et al. Toward a Comprehensive Atlas of the Physical Interactome of Saccharomyces cerevisiae. Molecular & Cellular Proteomics. 2007; 6(3):439–450. https://doi.org/10.1074/mcp.M600381-MCP200 PMID: 17200106
7. Ghoniem M, Fekete JD, Castagliola P. On the Readability of Graphs Using Node-link and Matrix-based Representations: A Controlled Experiment and Statistical Analysis. Information Visualization. 2005; 4 (2):114–135. https://doi.org/10.1057/palgrave.ivs.9500092
8. Behrisch M, Bach B, Henry Riche N, Schreck T, Fekete JD. Matrix Reordering Methods for Table and Network Visualization. Computer Graphics Forum. 2016; 35(3):693–716. https://doi.org/10.1111/cgf.12935
9. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An Information Aesthetic for Comparative Genomics. Genome Research. 2009; 19(9):1639–1645. https://doi.org/10.1101/gr.092759.109 PMID: 19541911
10. Kruskal JB, Landwehr JM. Icicle Plots: Better Displays for Hierarchical Clustering. The American Statistician. 1983; 37(2):162. https://doi.org/10.2307/2685881
11. Andrews K, Heidegger H. Information Slices: Visualising and Exploring Large Hierarchies Using Cascading, Semicircular Discs. In: InfoVis'98: Proc. IEEE Information Visualization Symposium, Carolina, USA; 1998. p. 9–12.
12. Stasko J, Zhang E. Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations. In: Proceedings of the IEEE Symposium on Information Vizualization (InfoVis '00). IEEE Computer Society Press; 2000. p. 57–65.
13. Johnson B, Shneiderman B. Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In: Proceedings of the IEEE Conference on Visualization (Vis '91); 1991. p. 284–291.
14. van Wijk JJ, van de Wetering H. Cushion Treemaps: Visualization of Hierarchical Information. In: In Proceedings of the IEEE Symposium on Information Visualization (InfoVis99); 1999. p. 73–78.
15. Soni U, Lu Y, Hansen B, Purchase HC, Kobourov S, Maciejewski R. The Perception of Graph Properties in Graph Layouts. Computer Graphics Forum. 2018; 37(3):169–181.
16. Borner K. Atlas of science: Visualizing what we know. Black hole of knowledge http://twitter.com/robinhanson/status/977261256443822080 The MIT Press, 2010
17. Ganglberger F, Kaczanowska J, Penninger JM, Hess A, Buehler K, Haubensak W. Predicting functional neuroanatomical maps from fusing brain networks with genetic information. bioRxiv. 2016; https://doi.org/10.1101/070037
18. Oh SW, Harris JA, Ng L, Winslow B, Cain N, Mihalas S, et al. A mesoscale connectome of the mouse brain. Nature. 2014; 508:207–214. https://doi.org/10.1038/nature13186 PMID: 24695228
19. Wenskovitch JE, Harris LA, Tapia JJ, Faeder JR, Marai GE. MOSBIE: a tool for comparison and analysis of rule-based biochemical models. BMC Bioinformatics. 2014; 15(1):316. https://doi.org/10.1186/1471-2105-15-316 PMID: 25253680
20. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, et al. Visualization of omics data for systems biology. Nature methods. 2010; 7:S56–S68. https://doi.org/10.1038/nmeth.1436 PMID: 20195258
21. Kuhner S. et al. Proteome organization in a genome-reduced bacterium. Science 326, 1235–1240, 2009. https://doi.org/10.1126/science.1176343 PMID: 19965468
22. Khaled ML, Bykhovskaya Y, Yablonski SER, Li H, Drewry MD, Aboobakar IF, et al. Differential Expression of Coding and Long Noncoding RNAs in Keratoconus-Affected Corneas. Investigative Ophthalmology & Visual Science. 2018; 59(7):2717–2728. https://doi.org/10.1167/iovs.18-24267 PMID: 29860458
23. Stone MC. A Field Guide to Digital Color. A K Peters/CRC Press; 2003.
24. Livingstone M. Vision and Art: The Biology of Seeing. Harry N. Abrams; 2008.
25. Luckiesh M. On retiring and advancing colors. American Journal of Psychology. 1918; 29(2):182–186.
26. Morris JH, Meng EC, Ferrin TE. Computational tools for the interactive exploration of proteomic and structural data. Mol Cell Proteomics. 2010; 9(8):1703–1715. https://doi.org/10.1074/mcp.R000007-MCP201 PMID: 20525940
27. Tufte ER. Envisioning Information. Envisioning Information. Graphics Press; 1990.

**28.** Preston CC, Wyles S, Reyes S, Storm EC, Eckloff BW, Faustino R. NUP155 insufficiency recalibrates a pluripotent transcriptome with network remodeling of a cardiogenic signaling module. BMC Systems Biology. 2018; 12(1). https://doi.org/10.1186/s12918-018-0590-x PMID: 29848314

**29.** Zhu M, Wang Q, Zhou W, Liu T, Yang L, Zheng P, et al. Integrated analysis of hepatic mRNA and miRNA profiles identified molecular networks and potential biomarkers of NAFLD. Scientific Reports. 2018; 8(7628). https://doi.org/10.1038/s41598-018-25743-8 PMID: 29769539

**30.** Shneiderman B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: Proc. of the IEEE Symp. on Visual Languages. IEEE Computer Society Press; 1996. p. 336–343.

**31.** Van Ham F, Perer A. "Search, show context, expand on demand": Supporting large graph exploration with degree-of-interest. IEEE Transactions on Visualization and Computer Graphics. 2009; 15(6).

**32.** Luciani T, Burks A, Sugiyama C, Komperda J, Marai GE. Details-First, Show Context, Overview Last: Supporting Exploration of Viscous Fingers in Large-Scale Ensemble Simulations. IEEE transactions on visualization and computer graphics. 2018; 25(1): 1225–1235.

**33.** Ware C, Franck G. Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions. ACM Transactions on Graphics. 1996; 15(2):121–139.

**34.** Andrei O., Fernandez M., Kirchner H., Pinaud B. Strategy-Driven Exploration for Rule-Based Models of Biochemical Systems with Porgy. Modeling Biomolecular Site Dynamics, Hlavacek B. (Eds), pp 43–70. Methods in Molecular Biology, vol. 1945 ( Springer). https://doi.org/10.1007/978-1-4939-9102-0_3

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for writing statistical book reviews

**Christopher J. Lortie**[1,2]*

**1** Department of Biology, York University, Toronto, ON, Canada, **2** The National Center for Ecological Analysis and Synthesis, UCSB, Santa Barbara, California, United States of America

* lortie@yorku.ca

## Abstract

Statistical books can provide deep insights into statistics and software. There are, however, many resources available to the practitioner. Book reviews have the capacity to function as a critical mechanism for the learner to assess the merits of engaging in part, in full, or at all with a book. The "ten simple rules" format, pioneered in computational biology, was applied here to writing effective book reviews for statistics because of the wide breadth of offerings in this domain, including topical introductions, computational solutions, and theory. Learning by doing is a popular paradigm in statistics and computation, but there is still a niche for books in the pedagogy of self-taught and instruction-based learning. Primarily, these rules ensure that book reviews function as a form of short syntheses to inform and guide readers in deciding to use a specific book relative to other options for resolving statistical challenges.





**Citation:** Lortie CJ (2019) Ten simple rules for writing statistical book reviews. PLoS Comput Biol 15(1): e1006562. https://doi.org/10.1371/journal.pcbi.1006562

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** January 24, 2019



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

### Author summary

Book reviews are a useful tool to inform learners in statistics and computational biology. As an ecologist, I teach biostatistics and use many resources in the analysis and coding of research data. In-depth texts can provide a critical resource, but well-written reviews can faciliate the decision to use a specific book.

## Introduction

Extensive resources now support the statistical programmer and analyst. The learner, reader, and general problem solver is thus faced with a choice of how to learn what is needed [1,2]. This brief synthesis is not intended to be a comment or criticism on the pedagogy associated with successfully acquiring statistical and coding expertise, but there is evidence suggesting that up to 80% of coders do not read books to learn how to code [6]. This seems like an unfortunate statistic, but the philosophy of "learning statistics by doing statistics" is not without merit and can be a viable approach to both introductory and expert learners alike [4]. Nonetheless, R, Python, SAS, and MATLAB/C++ are quite literally deep languages that need to be mastered. Fluency in a written or spoken language conveys reason and semantics [5]; statistical reasoning [4] with a corresponding representation of the associated mathematics [3] can likely be secured by both doing and reading [7]. Different problems and topics can also require the statistical programmer to embrace a diversity of resources to illuminate a solution, and the depth required must be defined by the prior knowledge of an individual and nature of the challenge.

Many statistical texts can be a significant time commitment, and open electronic resources are abundant. The decision to read a statistical programming book is not necessarily trivial. Short syntheses, i.e., a review, of the relative merits of a specific resource can provide a critical decision tool to the potential reader. The "ten simple simple rules" paper format was pioneered by Philip Bourne in *PLOS Computational Biology* [14], and it has proliferated to nearly 100 papers, all functioning as a succinct, unique form of synthesis in itself [8]. Sometimes extensive resources are summarized that support how to describe a focused process or get a task done in many domains of the scientific endeavor [11]. Of these "ten simple rules" papers, there have been three that address the review process, including how to be an effective referee [9], how to write a literature review [12], and how to write a reply paper [10]. Many of these rules certainly support improvements in how to write a review of statistical books and should be consulted. Yet, book reviews in the *Journal of Statistical Software*, e.g., strongly suggest that the importance of this topic warrants specific treatment because these reviews can serve many functions from descriptive summary to critical analysis to a launchpad for the importance of a statistical test, approach, program, language, and/or package. All are important functions that advance statistics, but at least some of the rules here can enhance their capacity to assess merit and need for the end practitioner. (Appropriately) defend books. Write reviews. Use reviews. Book reviews that effectively support the decision process for better statistical reasoning are needed. These rules promote this paradigm shift.

## Rules

### Rule 1: Introduce the topic

The book title is an excellent starting point for the reader to assess whether this is the resource for her but not the only mechanism. The book cover or sleeve synopsis and publisher description can also fail to capture the whole story, and some statistical treatises, both introductory and advanced, necessarily invoke related principles and topics [13]. As the objective expert of that specific text, an introduction to the necessity, scope, depth, and breadth of the topic in general can inform the reader on the challenges and solutions, including types of data or domains of inquiry that this field examines. Place the work within the span of the literature with a brief explanation of the area in which it is embedded. The goal of the first rule is therefore to ensure that the reader is in right place—conceptually, at least.

### Rule 2: State assumed audience (i.e., expertise-level for the reader)

Most technical book reviews state the level of expertise required by the reader. This is a critical form of synthesis that should be mentioned, even in brief, in a book review for statistics. The most typical categories range from introductory to advanced, with relatively higher-level offering described by "graduate student" and beyond as the reader. If the text is a blend of theory and practice with significant programming, the review should further explain the relative expertise needed for each and whether both dimensions are aligned in the assumed relative audience. Book reviews can also take the opportunity here to frame this assessment by the expertise of the referee (i.e., it is sometimes useful to know if the book reviewer is a statistician, a programmer, or a domain-specific end-user) or by the intended use of the text, such as primer, guide, in-depth treatise, or textbook appropriate for instruction at a given level.

## Rule 3: Explain different editions

If more than one edition exists, it is useful to describe the revisions to the more recent version of a book. Professional and teaching textbooks can be relatively expensive, and a critical assessment of value can provide instructors with the merits associated with potentially higher costs to students of purchasing a newer text. At minimum, a list of additions will facilitate a more informed choice for the reader and instructor, and mention of case studies, updates to code and data sets, and addition of supplements are all important criteria for the choice to learn or seek solutions from a specific edition.

## Rule 4: Describe the structure of chapters and associated pedagogy

Organization of the content matters to all learning [15], and content provides context [16]. The structure of statistical and programming texts can vary significantly. The length and complexity of chapters, use of headings and subsections within chapters, and end-of-chapter summaries are not always needed but often do no harm. Case studies, appendices, data sets, and location of supplements and supporting materials should be described. Contemporary texts in statistics are often a hybrid of print and electronic resource materials, and description of the extent that a text functions in this capacity can influence the decision by the reader based on her preferred modality of learning and the rationale for exploring this topic. This is also a good place to mention the different formats of the book (if available in print and online). As the reviewer, use parsimony and caution in deciding what level of detail to describe for the structural elements of a book—focus only on those elements that promoted the most effective learning.

## Rule 5: Summarize content

These are the results, so to speak, similar to a conventional scientific publication or study report. The description should be brief, topological, and highlight the most substantive elements of the book. This component of the book review need not be unduly critical but should provide an overview of the what the text offers. Some reviews take this description of what is done to also highlight what is done best and list the most valuable chapters to the reader. This can be a useful guide to the potential reader and a means to assess expectations from the book as a whole. If there are data sets or case studies that are revisited throughout the book or across multiple chapters, the extent that the chapters connect to one another can also be summarized. Mention whether the content of the book is serialized or if chapters can be read piecemeal.

## Rule 6: Critique readability

Readability is an intuitive concept. It is the ease that one can comprehend writing [17, 18]. Complexity in syntax, vocabulary, and sentence structure should be described in a review of a statistical book. A technical book need not be a technical challenge to read. More broadly, appeal, style, and interest are important to all but the most committed readers, and it is reasonable to assume that a book on statistics provide some sense of enthusiasm for the topic, compel the reader to think deeply, and engage one with the challenges explored. Composition is critical, particularly in long-form writing endeavors.

## Rule 7: List packages and linkages to statistical concepts

Within the R statistics community, there are now over 11,000 packages that extend the base language archived on https://cran.r-project.org. SAS Procs and libraries in Python and MATLAB are also extensive. Some statistical texts are associated with not only a single statistical program or language but with a single package or library. A review of a statistical book should thus describe the specificity of the book, explain the extent that the book relies on single solution sets, or conversely contrasts alternatives in different languages, applications, packages, and/or libraries, and frame the programming (if provided) to general statistical theory and reasoning. At times, this can be self-evident if the title of the book includes mention of the programming language or software, but the breadth of the statistics and case studies illustrated is typically not evident without review of the book. If the book is not tied to a specific computation tool in any form, then the reviewer should mention that this is the case and state that the concepts described can be applied and transferred broadly from the book.

## Rule 8: Compare the book to other resources

Compare and contrast. There is a wealth of both short- and long-form documentation available for many open coding languages used in statistics and data wrangling. There is also an extensive opportunity to seek specific solutions through numerous forums such as Stack Overflow (https://stackoverflow.com/questions/tagged/statistics), Cross Validated (https://stats.stackexchange.com), and Stack Exchange Mathematics (https://math.stackexchange.com). Online tutorials, blogs, carpentries, massive online open courses (MOOCs), and webinars often provide useful, and at times, deep-learning opportunities. A book review will certainly not comprehensively list all these options and compare and contrast to the principal subject text discussed, but if there is a significant alternative to consider, it should be mentioned. Finally, there are also other books. The reviewer should explicitly state the extent that she is contrasting to other resources, and due diligence by the reviewer suggests at the minimum a mention of the relative novelty and niche of the text in question.

## Rule 9: Comment on reading commitment and style

Reading a book is a relationship. The content, style, and perspective of the author(s) becomes a shared, internalized form of knowledge in a good book. As the reviewer, it is legitimate and useful to others to mention the extent that one enjoyed the text, connected with the writing and concepts, or struggled with certain elements (i.e., comment on the quality of the relationship with the book). A review should also mention the time that the reader should set aside to read and/or fully digest the content. If the "summarize content" rule proposed above did not mention the standout, best chapters, this is an excellent spot to describe the chapters that provided the most for your buck and should not be skipped. This is also an ideal opportunity to consider describing whether this is was a read-the-book-straight-through or piecemeal technical read for critical needs.

## Rule 10: Be professionally critical and state personal purpose

In general, it is best to be decisive in writing reviews [9]. Evaluate the capacity that the book delivers on its stated goal. Accept that you are part of the review process and likely have your own, specific purpose in reading this text. Admit this in the review by articulating the need, success of text, and decision (or not) to use the described tools, framework, or theory. Being specific and listing criteria point-by-point is useful to editors, authors, and readers [9]. Similar to the peer review process for papers, be balanced, fair, and professionally critical by mentioning both strengths and weaknesses from your perspective. Do your best to reveal implicit biases in your review.

## Summary

Reading, writing, and statistics. By putting oneself on the hook for a book to take notes and annotate or further synthesize these efforts and provide a review profoundly changes how one approaches a statistical and programming text [19, 20]. Higher education in the sciences and statistics has largely done away with book reviews and/or reports, but application and dissemination of critical thinking in statistics in the form of reviews is a learning opportunity. Capitalize on this process, particularly when using a text to solve a problem and write a review. Reviewing is a both a collegial and educational service that includes oneself as the beneficiary. The rules proposed herein for writing a book review for statistics and increasingly for the associated coding or implementation of statistics and data do not mean to imply that reading texts in this domain is a burden. On the contrary, the gratification of immersion in the structured reasoning inherent in these fields is a powerful form of literacy that merits discussion by people, for people. Recommendation algorithms certainly influence many aspects of human behavior, and a book review is a reminder to take a moment and savor the story.

## Acknowledgments

The York University Faculty Association and The National Center for Ecological Analysis and Synthesis supported this synthesis project.

## References

1. Ertmer P, Newby T. The expert learner: Strategic, self-regulated, and reflective. Instructional Science. 1996; 24: 1–24.
2. Mercer S. The beliefs of two expert EFL learners. The Language Learning Journal. 2011; 39: 57–74.
3. Widakdo WA. Mathematical Representation Ability by Using Project Based Learning on the Topic of Statistics. Journal of Physics: Conference Series. 2017; 895(1): 12–55.
4. Smith G. Learning Statistics by Doing Statistics. Journal of Statistics Education. 1998; 6(3): null.
5. Mol SE, Bus AG, de Jong MT. Interactive Book Reading in Early Education: A Tool to Stimulate Print Knowledge as Well as Oral Language. Review of Educational Research. 2009; 79(2): 979–1007.
6. DeMarco T Lister TR. Peopleware: productive projects and teams. London: Dorset House Publishing Company, Incorporated;1999.
7. Utts J, Sommer B, Acredolo C, Maher MW, Matthews HR. A Study Comparing Traditional and Hybrid Internet-Based Instruction in Introductory Statistics Classes. Journal of Statistics Education. 2003; 11 (3): null.
8. Dashnow H, Lonsdale A, Bourne PE. Ten Simple Rules for Writing a PLOS Ten Simple Rules Article. PLoS Comput Biol. 2014; 10(10): e1003858. https://doi.org/10.1371/journal.pcbi.1003858 PMID: 25340653
9. Bourne PE, Korngreen A. Ten simple rules for reviewers. PLoS Comput Biol. 2006; 2(9): 973–974.
10. Simmons MP. Ten Simple Rules for Writing a Reply Paper. PLoS Comput Biol. 2015; 11(10): e1004536. https://doi.org/10.1371/journal.pcbi.1004536 PMID: 26448197
11. Lortie CJ. Ten simple rules for short and swift presentations. PLoS Comput Biol. 2017; 13(3): e1005373. https://doi.org/10.1371/journal.pcbi.1005373 PMID: 28358832
12. Pautasso M. Ten Simple Rules for Writing a Literature Review. PLoS Comput Biol. 2013; 9(7): 1553–7358.
13. Gillespie C, Lovelace R. Efficient R programming. Sebastopol, California: Reilly Media; 2017.
14. Bourne PE. Ten simple rules for getting published. PLoS Comput Biol. 2005; 1(5): 341–342.
15. Blaich C, Wise K, Pascarella ET, Roksa J. Instructional Clarity and Organization: It's Not New or Fancy, But It Matters. Change: The Magazine of Higher Learning. 2016; 48(4): 6–13.
16. Grossman PL, Stodolsky SS. Content as Context: The Role of School Subjects in Secondary School Teaching. Educational Researcher. 1995; 24(8): 5–23.
17. Rudolph F. A new readability yardstick. Journal of Applied Psychology. 1948; 32: 221–223. PMID: 18867058
18. Dale E, Chall JS. The Concept of Readability. Elementary English. 1949; 26(1): 19–26.
19. Balajthy E. Using Student-Constructed Questions to Encourage Active Reading. Journal of Reading. 1984; 27(5): 408–411.
20. Tashman C, Edwards K. Active reading and its discontents: the situations, problems and ideas of readers. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011;2927–2936.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for collaboratively writing a multi-authored paper

**Marieke A. Frassl[1]*‡, David P. Hamilton[1], Blaize A. Denfeld[2], Elvira de Eyto[3], Stephanie E. Hampton[4], Philipp S. Keller[5], Sapna Sharma[6], Abigail S. L. Lewis[7], Gesa A. Weyhenmeyer[8], Catherine M. O'Reilly[9], Mary E. Lofton[10], Núria Catalán[11]**

**1** Australian Rivers Institute, Griffith University, Nathan, Queensland, Australia, **2** Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden, **3** Marine Institute, Furnace, Newport, Co. Mayo, Ireland, **4** Center for Environmental Research, Education and Outreach, Washington State University, Pullman, Washington, United States of America, **5** UFZ, Helmholtz Centre for Environmental Research, Department of Lake Research, Magdeburg, Germany, **6** Department of Biology, York University, Toronto, Ontario, Canada, **7** Department of Biology, Pomona College, Claremont, California, United States of America, **8** Department of Ecology and Genetics/Limnology, Uppsala University, Uppsala, Sweden, **9** Department of Geography, Geology, and the Environment, Illinois State University, Normal, Illinois, United States of America, **10** Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, United States of America, **11** Catalan Institute for Water Research (ICRA), Girona, Spain

‡ MAF is the lead author. All authors contributed equally to this work. Besides for MAF, author order was computed randomly.
* m.frassl@griffith.edu.au





**Citation:** Frassl MA, Hamilton DP, Denfeld BA, de Eyto E, Hampton SE, Keller PS, et al. (2018) Ten simple rules for collaboratively writing a multi-authored paper. PLoS Comput Biol 14(11): e1006508. https://doi.org/10.1371/journal.pcbi.1006508

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** November 15, 2018



**Funding:** This work was supported by the Global Lake Ecological Observatory Network (GLEON; www.gleon.org). ML and PK received the GLEON student travel fund. GW was supported by the Swedish Research Council, Grant No. 2016-04153. NC had the support of the Beatriu de Pinós postdoctoral programme (BP-2016-00215). PK was supported by the DFG (Deutsche Forschungsgemeinschaft, grant SP 1570/1-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Science is increasingly done in large teams [1], making it more likely that papers will be written by several authors from different institutes, disciplines, and cultural backgrounds. A small number of "Ten simple rules" papers have been written on collaboration [2, 3] and on writing [4, 5] but not on combining the two. Collaborative writing with multiple authors has additional challenges, including varied levels of engagement of coauthors, provision of fair credit through authorship or acknowledgements, acceptance of a diversity of work styles, and the need for clear communication. Miscommunication, a lack of leadership, and inappropriate tools or writing approaches can lead to frustration, delay of publication, or even the termination of a project.

To provide insight into collaborative writing, we use our experience from the Global Lake Ecological Observatory Network (GLEON) [6] to frame 10 simple rules for collaboratively writing a multi-authored paper. We consider a collaborative multi-authored paper to have three or more people from at least two different institutions. A multi-authored paper can be a result of a single discrete research project or the outcome of a larger research program that includes other papers based on common data or methods. The writing of a multi-authored paper is embedded within a broader context of planning and collaboration among team members. Our recommended rules include elements of both the planning and writing of a paper, and they can be iterative, although we have listed them in numerical order. It will help to revisit the rules frequently throughout the writing process. With the 10 rules outlined below, we aim to provide a foundation for writing multi-authored papers and conducting exciting and influential science.

## Rule 1: Build your writing team wisely

The writing team is formed at the beginning of the writing process. This can happen at different stages of a research project. Your writing team should be built upon the expertise and interest of your coauthors. A good way to start is to review the initial goal of the research project and to gather everyone's expectations for the paper, allowing all team members to decide whether they want to be involved in the writing. This step is normally initiated by the research project leader(s). When appointing the writing team, ensure that the team has the collective expertise required to write the paper and stay open to bringing in new people if required. If you need to add a coauthor at a later stage, discuss this first with the team (Rule 8) and be clear as to how the person can contribute to the paper and qualify as a coauthor (Rules 4 and 10). When in doubt about selecting coauthors, in general we suggest to opt for being inclusive. A shared list with contact information and the contribution of all active coauthors is useful for keeping track of who is involved throughout the writing process.

In order to share the workload and increase the involvement of all coauthors during the writing process, you can distribute specific roles within the team (e.g., a team leader and a facilitator [see Rule 2] and a note taker [see Rule 8]).

## Rule 2: If you take the lead, provide leadership

Leadership is critical for a multi-authored paper to be written in a timely and satisfactory manner. This is especially true for large, joint projects. The leader of the writing process and first author typically are the same person, but they don't have to be. The leader is the contact person for the group, keeps the writing moving forward, and generally should manage the writing process through to publication. It is key that the leader provides strong communication and feedback and acknowledges contributions from the group. The leader should incorporate flexibility with respect to timelines and group decisions. For different leadership styles, refer to [7, 8].

When developing collaborative multi-authored papers, the leader should allow time for all voices to be heard. In general, we recommend leading multi-authored papers through consensus building and not hierarchically because the manuscript should represent the views of all authors (Rule 9). At the same time, the leader needs to be able to make difficult decisions about manuscript structure, content, and author contributions by maintaining oversight of the project as a whole.

Finally, a good leader must know when to delegate tasks and share the workload, e.g., by delegating facilitators for a meeting or assigning responsibilities and subleaders for sections of a manuscript. At times, this may include recognizing that something has changed, e.g., a change in work commitments by a coauthor or a shift in the paper's focus. In such a case, it may be timely for someone else to step in as leader and possibly also as first author, while the previous leader's work is acknowledged in the manuscript or as a coauthor (Rule 4).

## Rule 3: Create a data management plan

If not already implemented at the start of the research project, we recommend that you implement a data management plan (DMP) that is circulated at an early stage of the writing process and agreed upon by all coauthors (see also [9] and https://dmptool.org/; https://dmponline.dcc.ac.uk/). The DMP should outline how project data will be shared, versioned, stored, and curated and also details of who within the team will have access to the (raw) data during and post publication.

Multi-authored papers often use and/or produce large datasets originating from a variety of sources or data contributors. Each of these sources may have different demands about how data and code are used and shared during analysis and writing and after publication. Previous articles published in the "Ten simple rules" series provide guidance on the ethics of big-data research [10], how to enable multi-site collaborations through open data sharing [3], how to store data [11], and how to curate data [12]. As many journals now require datasets to be shared through an open access platform as a prerequisite to paper publication, the DMP should include detail on how this will be achieved and what data (including metadata) will be included in the final dataset.

Your DMP should not be a complicated, detailed document and can often be summarized in a couple of paragraphs. Once your DMP is finalized, all data providers and coauthors should confirm that they agree with the plan and that their institutional and/or funding agency obligations are met. It is our experience within GLEON that these obligations vary widely across the research community, particularly at an intercontinental scale.

## Rule 4: Jointly decide on authorship guidelines

Defining authorship and author order are longstanding issues in science [13]. In order to avoid conflict, you should be clear early on in the research project what level of participation is required for authorship. You can do this by creating a set of guidelines to define the contributions and tasks worthy of authorship. For an authorship policy template, see [14] and check your institute's and the journal's authorship guidelines. For example, generating ideas, funding acquisition, data collection or provision, analyses, drafting figures and tables, and writing sections of text are discrete tasks that can constitute contributions for authorship (see, e.g., the CRediT system: http://docs.casrai.org/CRediT [15]). All authors are expected to participate in multiple tasks, in addition to editing and approving the final document. It is debated whether merely providing data does qualify for coauthorship. If data provision is not felt to be grounds for coauthorship, you should acknowledge the data provider in the Acknowledgments [16].

Your authorship guidelines can also increase transparency and help to clarify author order. If coauthors have contributed to the paper at different levels, task-tracking and indicating author activity on various tasks can help establish author order, with the person who contributed most in the front. Other options include groupings based on level of activity [17] or having the core group in the front and all other authors listed alphabetically. If every coauthor contributed equally, you can use alphabetical order [18] or randomly assigned order [19]. Joint first authorship should be considered when appropriate. We encourage you to make a statement about author order (e.g., [19]) and to generate authorship attribution statements; many journals will include these as part of the Acknowledgments if a separate statement is not formally required. For those who do not meet expectations for authorship, an alternative to authorship is to list contributors in the Acknowledgments [15]. Be aware of coauthors' expectations and disciplinary, cultural, and other norms in what constitutes author order. For example, in some disciplines, the last author is used to indicate the academic advisor or team leader. We recommend revisiting definitions of authorship and author order frequently because roles and responsibilities may change during the writing process.

## Rule 5: Decide on a writing strategy

The writing strategy should be adapted according to the needs of the team (white shapes in Fig 1) and based on the framework given through external factors (gray shapes in Fig 1). For example, a research paper that uses wide-ranging data might have several coauthors but one principal writer (e.g., a PhD candidate) who was conducting the analysis, whereas a comment or review in a specific research field might be written jointly by all coauthors based on parallel discussion. In most cases, the approach that everyone writes on everything is not possible and is very inefficient. Most commonly, the paper is split into sub-sections based on what aspects of the research the coauthors have been responsible for or based on expertise and interest of the coauthors. Regardless of which writing strategy you choose, the importance of engaging all

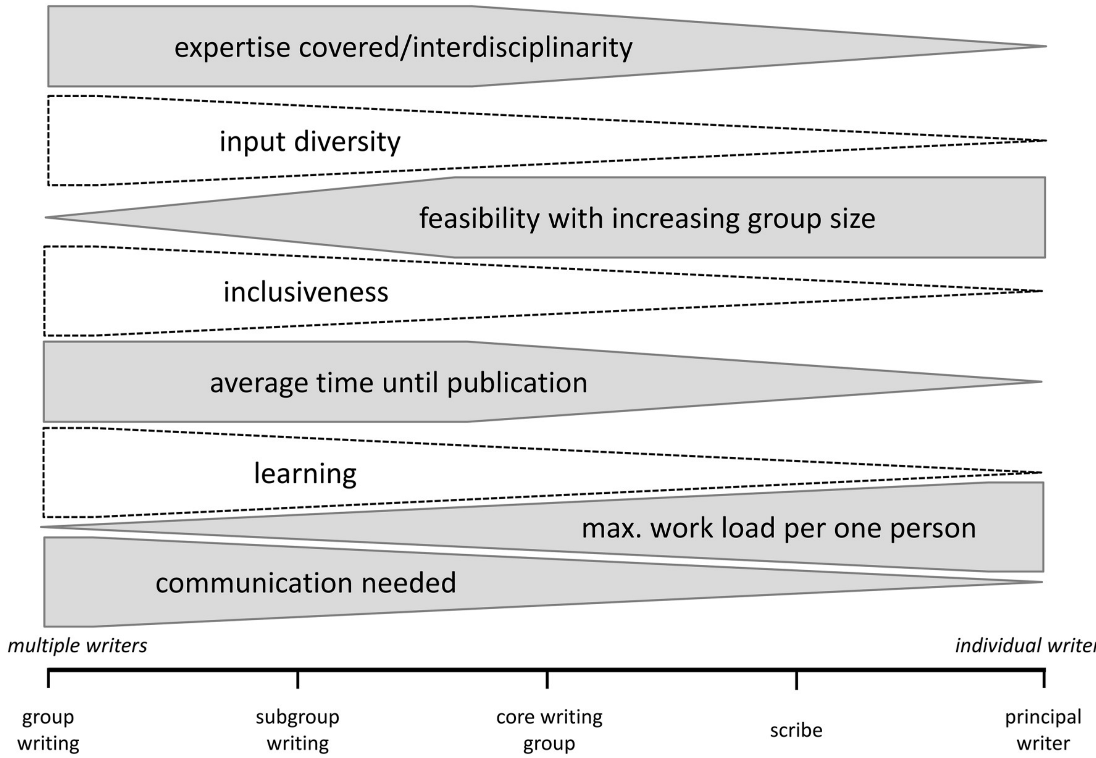



**Fig 1. Decision chart for writing strategy.** Different writing strategies ranging from very inclusive to minimally inclusive: group writing = everyone writes on everything; subgroup writing = document is split up into expertise areas, each individual contributes to a subsection; core writing group = a subgroup of a few coauthors writes the paper; scribe writing = one person writes based on previous group discussions; principal writer = one person drafts and writes the paper (writing styles adapted from [20]). Which writing strategy you choose depends on external factors (filled, gray shapes), such as the interdisciplinarity of the study or the time pressure of the paper to be published, and affects the payback (dashed, white shapes). An increasing height of the shape indicates an increasing quantity of the decision criteria, such as the interdisciplinarity, diversity, feasibility, etc.

https://doi.org/10.1371/journal.pcbi.1006508.g001

team members in defining the narrative, format, and structure of the paper cannot be overstated; this will preempt having to rewrite or delete sections later.

For an efficient writing process, try to use the active voice in suggestions and make direct edits rather than simply stating that a section needs revision. For all writing strategies, the lead author(s) has to ensure that the completed text is cohesive.

## Rule 6: Choose digital tools to suit your needs

A suitable technology for writing your multi-authored paper depends upon your chosen writing approach (Rule 5). For projects in which the whole group writes together, synchronous technologies such as Google Docs or Overleaf work well by allowing for interactive writing that facilitates version control (see also [21]). In contrast, papers written sequentially, in parallel by subsections, or by only one author may allow for using conventional programs such as Microsoft Word or LibreOffice. In any case, you should create a plan early on for version control, comments, and tracking changes. Regularly mark the version of the document, e.g., by including the current date in the file name. When working offline and distributing the document, add initials in the file name to indicate the progress and most recent editor.

High-quality communication is important for efficient discussion on the paper's content. When picking a virtual meeting technology, consider the number of participants permitted in a single group call, ability to record the meeting, audio and visual quality, and the need for additional features such as screencasting or real-time notes. Especially for large groups, it can be helpful for people who are not currently speaking to mute their microphones (blocking background noise), to use the video for nonverbal communication (e.g., to show approval or rejection and to help nonnative speakers), or to switch off the video when internet speeds are slow. More guidelines for effective virtual meetings are available in Hampton and colleagues [22].

In between virtual meetings, virtual technologies can help to streamline communication (e.g., https://slack.com) and can facilitate the writing process through shared to-do lists and task boards including calendar features (e.g., http://trello.com).

With all technologies, accessibility, ease of use, and cost are important decision criteria. Note that some coauthors will be very comfortable with new technologies, whereas others may not be. Both should be ready to compromise in order to be as efficient and inclusive as possible. Basic training in unfamiliar technologies will likely pay off in the long term.

## Rule 7: Set clear timelines and adhere to them

As for the overall research project, setting realistic and effective deadlines maintains the group's momentum and facilitates on-schedule paper completion [23]. Before deciding to become a coauthor, consider your own time commitments. As a coauthor, commit to set deadlines, recognize the importance of meeting them, and notify the group early on if you realize that you will not be able to meet a deadline or attend a meeting. Building consensus around deadlines will ensure that internally imposed deadlines are reasonably timed [23] and will increase the likelihood that they are met. Keeping to deadlines and staying on task require developing a positive culture of encouragement within the team [14]. You should respect people's time by being punctual for meetings, sending out drafts and the meeting agenda on schedule, and ending meetings on time.

To develop a timeline, we recommend starting by defining the "final" deadline. Occasionally, this date will be set "externally" (e.g., by an editorial request), but in most cases, you can set an internal consensus deadline. Thereafter, define intermediate milestones with clearly defined tasks and the time required to fulfill them. Look for and prioritize strategies that allow multiple tasks to be completed simultaneously because this allows for a more efficient timeline. Keep in mind that "however long you give yourself to complete a task is how long it will take" [24] and that group scheduling will vary depending on the selected writing strategy (Rule 5). Generally, collaborative manuscripts need more draft and revision rounds than a "solo" article.

## Rule 8: Be transparent throughout the process

This rule is important for the overall research project but becomes especially important when it comes to publishing and coauthorship. Being as open as possible about deadlines (Rule 7) and expectations (including authorship, Rule 4) helps to avoid misunderstandings and conflict. Be clear about the consequences if someone does not follow the group's rules but also be open to rediscuss rules if needed. Potential consequences of not following the group's rules include a change in author order or removing authorship. It should also be clear that a coauthor's edits might not be included in the final text if s/he does not contribute on time. Bad experience from past collaboration can lead to exclusion from further research projects.

As for collaboration [2], communication is key. During meetings, decide on a note taker who keeps track of the group's discussions and decisions in meeting notes. This will help coauthors who could not attend the meeting as well as help the whole group follow up on decisions later on. Encourage everyone to provide feedback and be sincere and clear if something is not working—writing a multi-authored paper is a learning process. If you feel someone is frustrated, try to address the issue promptly within the group rather than waiting and letting the problem escalate. When resolving a conflict, it is important to actively listen and focus the conversation on how to reach a solution that benefits the group as a whole [25]. Democratic decisions can often help to resolve differing opinions.

## Rule 9: Cultivate equity, diversity, and inclusion

Multi-authored papers will likely have a team of coauthors with diverse demographics and cultural values, which usually broadens the scope of knowledge, experience, and background. While the benefit of a diverse team is clear [14], successfully integrating diversity in a collaborative team effort requires increased awareness of differences and proactive conflict management [25]. You can cultivate diversity by holding members accountable to equity, diversity, and inclusivity guidelines (e.g., https://www.ryerson.ca/edistem/).

If working across cultures, you will need to select the working language (both for verbal and written communications); this is most commonly the publication language. When team members are not native speakers in the working language, you should always speak slowly, enunciate clearly, and avoid local expressions and acronyms, as well as listen closely and ask questions if you do not understand. Besides language, be empathetic when listening to others' opinions in order to genuinely understand your coauthors' points of view [26].

When giving verbal or written feedback, be constructive but also be aware of how different cultures receive and react to feedback [27]. Inclusive writing and speaking provide engagement, e.g., "*we* could do that," and acknowledge input between peers. In addition, you can create opportunities for expression of different personalities and opinions by adopting a participatory group model (e.g., [28]).

## Rule 10: Consider the ethical implications of your coauthorship

Being a coauthor is both a benefit and a responsibility: having your name on a publication implies that you have contributed substantially, that you are familiar with the content of the paper, and that you have checked the accuracy of the content as best you can. To conduct a self-assessment as to whether your contributions merit coauthorship, start by revisiting authorship guidelines for your group (Rule 4).

Be sure to verify the scientific accuracy of your contributions; e.g., if you contributed data, it is your responsibility that the data are correct, or if you performed laboratory or data analyses, it is your responsibility that the analyses are correct. If an author is accused of scientific misconduct, there are likely to be consequences for all the coauthors. Although there are currently no clear rules for coauthor responsibility [29], be aware of your responsibility and find a balance between trust and control.

One of the final steps before submission of a multi-authored paper is for all coauthors to confirm that they have contributed to the paper, agree upon the final text, and support its submission. This final confirmation, initiated by the lead author, will ensure that all coauthors have considered their role in the work and can affirm contributions. It is important that you repeat the confirmation step each time the paper is revised and resubmitted. Set deadlines for the confirmation steps and make clear that coauthorship cannot be guaranteed if confirmations are not done.

## Conclusion

When writing collaborative multi-authored papers, communication is more complex, and consensus can be more difficult to achieve. Our experience shows that structured approaches can help to promote optimal solutions and resolve problems around authorship as well as data ownership and curation. Clear structures are vital to establish a safe and positive environment that generates trust and confidence among the coauthors [14]. The latter is especially challenging when collaborating over large distances and not meeting face-to-face.

Since there is no single "right approach," our rules can serve as a starting point that can be modified specifically to your own team and project needs. You should revisit these rules frequently and progressively adapt what works best for your team and the project.

We believe that the benefits of working in diverse groups outweigh the transaction costs of coordinating many people, resulting in greater diversity of approaches, novel scientific outputs, and ultimately better papers. If you bring curiosity, patience, and openness to team science projects and act with consideration and empathy, especially when writing, the experience will be fun, productive, and rewarding.

## Acknowledgments

We would like to thank Meredith Holgerson and Samantha Oliver for their input in the very beginning of this project. We thank the Global Lake Observatory Network (GLEON), which has provided a trustworthy and collaborative work environment.

## References

1. Wuchty S, Jones BF, Uzzi B. The Increasing Dominance of Teams in Production of Knowledge. Science. 2007; 316:1036–1039. https://doi.org/10.1126/science.1136099 PMID: 17431139
2. Vicens Q, Bourne PE. Ten Simple Rules for a Successful Collaboration. PLoS Comput Biol. 2007; 3(3): e44. https://doi.org/10.1371/journal.pcbi.0030044 PMID: 17397252
3. Boland MR, Karczewski KJ, Tatonetti NP. Ten Simple Rules to Enable Multi-site Collaborations through Data Sharing. PLoS Comput Biol. 2017; 13(1):e1005278. https://doi.org/10.1371/journal.pcbi.1005278 PMID: 28103227
4. Mensh B, Kording K. Ten simple rules for structuring papers. PLoS Comput Biol. 2017; 13(9): e1005619. https://doi.org/10.1371/journal.pcbi.1005619 PMID: 28957311
5. Weinberger CJ, Evans JA, Allesina S. Ten Simple (Empirical) Rules for Writing Science. PLoS Comput Biol. 2015; 11(4):e1004205. https://doi.org/10.1371/journal.pcbi.1004205 PMID: 25928031
6. Weathers K, Hanson PC, Arzberger P, Brentrup J, Brookes JD, Carey CC, et al. The Global Lake Ecological Observatory Network (GLEON): the evolution of grassroots network science. Limnol Oceanogr Bulletin. 2013; 22(3): 71–73.
7. Bass BM, Avolio BJ. Improving Organizational Effectiveness Through Transformational Leadership. 1st ed. London: SAGE Publications; 1994.
8. Braun S, Peus C, Frey D, Knipfer K. Leadership in Academia: Individual and Collective Approaches to the Quest for Creativity and Innovation. In: Peus C, Braun S, Schyns B, editors. Leadership Lessons from Compelling Contexts: Emerald Group Publishing Limited; 2016. pp. 349–365.
9. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol. 2015; 11(10):e1004525. https://doi.org/10.1371/journal.pcbi.1004525 PMID: 26492633
10. Zook M, Barocas S, Crawford K, Keller E, Gangadharan SP, Goodman A, et al. Ten simple rules for responsible big data research. PLoS Comput Biol. 2017; 13(3):e1005399. https://doi.org/10.1371/journal.pcbi.1005399 PMID: 28358831
11. Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. Ten Simple Rules for Digital Data Storage. PLoS Comput Biol. 2016; 12(10):e1005097. https://doi.org/10.1371/journal.pcbi.1005097 PMID: 27764088
12. Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol. 2014; 10(4):e1003542. https://doi.org/10.1371/journal.pcbi.1003542 PMID: 24763340
13. Hunt R. Trying an authorship index. Nature. 1991; 352(6332):187–187.
14. Cheruvelil KS, Soranno PA, Weathers KC, Hanson PC, Goring SJ, Filstrup CT, et al. Creating and maintaining high-performing collaborative research teams: the importance of diversity and interpersonal skills. Front Ecol Environ. 2014; 12(1):31–38. https://doi.org/10.1890/130001

**15.** Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. Learn Publ. 2015; 28(2):151–155. https://doi.org/10.1087/20150211

**16.** Duke CS, Porter JH. The Ethics of Data Sharing and Reuse in Biology. Bioscience. 2013; 63(6):483–489. https://doi.org/10.1525/bio.2013.63.6.10

**17.** O'Reilly CM, Sharma S, Gray DK, Hampton SE, Read JS, Rowley RJ, et al. Rapid and highly variable warming of lake surface waters around the globe. Geophys Res Lett. 2015; 42(24): 10773–10781. https://doi.org/10.1002/2015GL066235

**18.** Pastor A, Lupon A, Gomez-Gener L, Rodriguez-Castillo T, Abril M, Arce MI, et al. Local and regional drivers of headwater streams metabolism: insights from the first AIL collaborative project. Limnetica. 2017; 36(1):67–85.

**19.** Gibson CA, O'Reilly CM. Organic matter stoichiometry influences nitrogen and phosphorus uptake in a headwater stream. Freshw Sci. 2012; 31(2):395–407. https://doi.org/10.1899/11-033.1

**20.** Noël S, Robert J-M. Empirical Study on Collaborative Writing: What Do Co-authors Do, Use, and Like? Comput Support Coop Work. 2004; 13(1):63–89. https://doi.org/10.1023/B:COSU.0000014876.96003.be

**21.** Hampton SE, Anderson SS, Bagby SC, Gries C, Han X, Hart EM, et al. The Tao of open science for ecology. Ecosphere. 2015; 6(7):1–13. https://doi.org/10.1890/ES14-00402.1

**22.** Hampton SE, Halpern BS, Winter M, Balch JK, Parker JN, Baron JS, et al. Best Practices for Virtual Participation in Meetings: Experiences from Synthesis Centers. The Bulletin of the Ecological Society of America. 2017; 98(1):57–63. https://doi.org/10.1002/bes2.1290

**23.** Ariely D, Wertenbroch K. Procrastination, Deadlines, and Performance: Self-Control by Precommitment. Psychol Sci. 2002; 13(3):219–224. https://doi.org/10.1111/1467-9280.00441 PMID: 12009041

**24.** Northcote Parkinson C. Parkinson's Law: Or the Pursuit of Progress. 1st ed. London: John Murray; 1958.

**25.** Bennett LM, Gadlin H, Levine-Finley S. Collaboration & Team Science: A Field Guide. National Institutes of Health. August 2010. https://ccrod.cancer.gov/confluence/display/NIHOMBUD/Home. [cited 19 February 2018].

**26.** Covey SR. The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change. London: Simon & Schuster; 2004.

**27.** Meyer E. The Culture Map (INTL ED): Decoding How People Think, Lead, and Get Things Done Across Cultures. 1st ed. New York: PublicAffairs; 2016.

**28.** Kaner S. Facilitator's guide to participatory decision-making. 1st ed. San Francisco: Jossey-Bass; 2014.

**29.** Andersen H. Co-author responsibility: Distinguishing between the moral and epistemic aspects of trust. EMBO reports. 2014; 15(9):914–918. https://doi.org/10.15252/embr.201439016 PMID: 25114099

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for scientists: Improving your writing productivity

**Todd C. Peterson[1], Sofie R. Kleppner[2], Crystal M. Botham[3]***

**1** Department of Neurology & Neurological Sciences, Stanford University, Stanford, California, United States of America, **2** Office of Postdoctoral Affairs, Stanford University, Stanford, California, United States of America, **3** Stanford Biosciences Grant Writing Academy, Stanford University, Stanford, California, United States of America

* cbotham@stanford.edu



OPEN ACCESS

**Citation:** Peterson TC, Kleppner SR, Botham CM (2018) Ten simple rules for scientists: Improving your writing productivity. PLoS Comput Biol 14 (10): e1006379. https://doi.org/10.1371/journal.pcbi.1006379

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** October 4, 2018



**Funding:** The authors received no funding for this manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

> . . . As a scientist, you are a professional writer. Writing is as important a tool in your toolbox as molecular biology, chemical analysis, statistics, or other purely "scientific" tools. Some of these tools allow us to generate data; others to analyze and communicate results. Writing is the most important of the latter. Because it forms the bridge to your audience, it can act as the rate-limiting step that constrains the effectiveness of all other tools.—Joshua Schimel [1]

Science requires communicating new and exciting findings to diverse audiences. Written communication is especially critical for our success as scientists because we must write to receive degrees (e.g., dissertations), share our discoveries (e.g., manuscripts and abstracts for professional meetings), request funding (e.g., grants, contracts), etc. The process of writing can also refine our research because good writing is an iterative process, with feedback leading to new ideas and experimental follow-up. In reality, we often postpone working on communication until we feel ready or a deadline is imminent, which diminishes our ability for clear writing. We often procrastinate sharing our writing because we consider our audience to be only the peer reviewers of grants and manuscripts. We neglect the opportunity to use perhaps the most important audience of all: ourselves. The best way to engage with ourselves is to develop a strong and sustained writing practice.

Writing every day, even for a short time, improves our thinking and our productivity as scientists. It provides time and space for reflection, allowing new ideas to mature, and maintains perspective on challenging work. We agree with Scott Montgomery that ". . . 'clear thinking can emerge from clear writing.' Imposing order by organizing and expressing ideas has great power to clarify. In many cases, writing is the process through which scientists come to understand the real form and implications of their work" [2]. However, even with the best of intentions, it is easy to postpone writing (Box 1). We believe that establishing a writing practice

### Box 1. Hurdles to writing that all scientists can face

Types of Writing Resistance

- Evaluation: "This draft stinks."
- Inspiration: "I don't have a good idea yet."

- Motivation: "I just don't feel like it."
- Optimization: "I need to make this sentence perfect."
- Procrastination: "I will start working on it tomorrow."
- Separation: "I need a lot of time in a quiet place to write."
- Temptation: "My lab bench is really disorganized; it needs to be cleaned now."

must be a deliberate act. We further believe that the pay-off for establishing this practice will be found in increased productivity and impact. Here, we outline 10 simple rules for improving your writing productivity, which will also enhance your thinking as a scientist.

## Rule 1: Define your writing time

> The essential key for writing is to write regularly—like it or not—great ideas come often by writing; releasing the subconscious—waiting for inspiration and ideas will not work, but it does help to have a notebook with you all the time for sudden brainstorms or inspiration.—Dr. Robert Marc Friedman [3]

Commit yourself to writing daily or at least three to four times per week by defining the time when you will write. Pick a bitesize chunk of time where you are unlikely to have conflicts, so that this time is protected for writing and nothing else. For example, if writing right before bed results in a conflict with exhaustion, choose a time earlier in the day. Setting aside just 15 to 30 minutes each day may be sufficient because even short amounts of time can enable meaningful increases in your productivity. As you set your initial goal, consider the SMART criteria: specific, measurable, action-oriented, reasonable, time-bound [4]. For example: "I am going to write from 8:30 AM to 9:30 AM on weekdays."

Write down your goal on a sticky note, and place it where you will have a daily reminder of your goal—attach it to your coffee grinder or stick it on your computer screen or near your pipettes. Block off the time on your calendar as another permanent reminder of your commitment to writing. You might also use a calendar reminder or an alarm on your phone to catch your attention when it is time to start writing. Then, when you are more comfortable with your writing schedule, consider building up the time you spend writing.

## Rule 2: Create a working environment that really works

Where are you most likely to settle in and write without external distractions? Some people work best in a café or on a flight. Others need complete silence and are most effective working in the library. Once you have identified your preferred writing space, learn to manage distractions. If you find yourself focused on that email you forgot to send, or if finding an antibody for your upcoming experiment suddenly seems urgent, you may want to start a running "to do" list and address it at the conclusion of your writing time. Resist the urge to write a quick email or other small task because it creates further opportunities to procrastinate. Unavoidable interruptions can be managed by saying, "I am committing this time to writing, can I contact you when I am done?" If you have difficulty ignoring the siren call of new emails or the internet, turn it off for the duration of your writing time. Constant interruptions disrupt your thinking and eat up your time for writing.

## Rule 3: Write first, edit later

During your designated writing time: Write! If you feel stuck, try writing whatever pops into your head for the first five minutes, or label your draft as a "cruddy first draft," or write your thoughts as a letter to someone who you care about so you are not distracted by making it perfect. Just putting something down on paper will drive the creative process. Ignore your internal editor when it interferes—your task is to write without inhibition. The evaluative component, i.e., editing and polishing, should come later.

## Rule 4: Use triggers to develop a productive writing habit

Maintaining a concrete habit can shift your attitude towards writing and reduce anxiety about how to start, how to finish, and how to make your work flow. To ensure your habit develops, make use of triggers to kick off that automatic urge to write like taking a brisk walk before writing or making a pot of your favorite tea. Maybe you decide to write after your department's weekly research seminar. Music can be an effective trigger too, as long as it doesn't take too much of your attention. Spend some time determining a few effective triggers and strategically place these before your writing time to help you write routinely. Your goal should be to make writing a habit, like brushing your teeth, so even if you don't feel like writing you still do it.

## Rule 5: Be accountable

Establishing any new habit is hard without accountability. How can you best ensure that you stick to it? You can work with rewards—allowing yourself to read a chapter in that new book you love, watch a short video, or have a second cup of coffee after completing your writing time. We have found it helpful to pair up with a writing buddy. For example, once you have a time established, you can text or email your colleague to say, "I'm starting," and then contact them again when you are done. The simple act of communicating your habit can help you keep on course. You could also arrange to meet somewhere and write together, as long as you remain focused on the tasks at hand.

## Rule 6: Seek feedback and ask for what you want

It can be scary to share your drafts and elicit feedback, but writing effectively is not a solitary activity. Eventually someone is going to read what you wrote, and you can help make that experience a good one for your audience. You can and should elicit feedback from all sorts of people. Ask for feedback early and often from fellow graduate students or postdocs, other colleagues or collaborators, and research mentors, both within and outside of your discipline. Nonscientists can also provide valuable feedback about the clarity of your ideas. When you share a draft, give your reader some idea of where you are in the writing process and what kinds of things you would like them to focus on. For example, you might ask whether the first paragraph is clear, or what they think the major point is, or if the overall argument is persuasive. Many readers end up focusing on grammar, so let them know ahead of time if you want something more than copyediting. Ask your readers to prioritize their feedback, and give you their top three issues in detail. Good writing requires this iterative process, and each revision will refine your writing.

## Rule 7: Think about what you're writing outside of your scheduled writing time

Sometimes the best thinking is done when you are otherwise occupied with a mundane task. Similarly, thinking about what you are working on outside your scheduled writing time can lead to more effective writing. Take time away from your writing to allow your mind to churn. Go to the animal facility to manage the murine colony, image cells on the confocal microscope, talk about your latest exciting result, etc. Use this time outside your scheduled writing time to mentally outline early drafts, deepen your argument, refine your hypothesis, etc. Most importantly, thinking about your writing can stimulate your desire to write.

## Rule 8: Practice, practice, practice

Just like any other skill, through practice you will become a better writer. Seek out opportunities to learn new approaches to be a more effective writer. For example, if you are currently working on a manuscript, learn more about that genre. As an example, see the Massive Open Online Course *Writing in the Sciences* [5] and these other sources for best practices for composing effective manuscripts [6–8]. Reading relentlessly will also help you to become a better writer, so reread those high-impact papers in your field and analyze the authors' effectiveness at communicating the significance of their research. Do not limit yourself to scientific papers; read a variety of good writing. It is also critical to sharpen your writing skills through exploring writing style and elegance. We recommend reading a classic book on the topic, *Elements of Style* [9]. It can be helpful to practice engaging audiences by writing for diverse genres (i.e., manuscript, research plan for a proposal, lay abstract, commentary, blog, etc.), which will also keep your writing practice from becoming monotonous.

We also recommend that you surround yourself with a supportive community of other writers. Learn from these writers by offering to read and provide feedback on their writing. Asking for feedback and providing it to others will make you a more skilled editor, mentor, and writer.

## Rule 9: Manage your self-talk about writing

Be mindful of your self-talk about writing because negative thinking is detrimental to your writing practice and can squelch your writing. Make a conscious effort to silence your internal editor and redirect demoralizing messages ("There is too much to do. I don't have enough time.") to positive, hopeful thoughts ("I am going to make progress today by writing for one hour"). Remind yourself that even incremental progress will lead to something bigger with time.

## Rule 10: Reevaluate your writing practice often

Track your writing, perhaps by marking it off on a calendar, which can serve as a reminder when you have missed a few days. It is normal for your writing practice to have cyclic ups and downs. Don't be discouraged; all writers must overcome writing resistance at some time or another. If your writing practice is not where you want it to be, think about what is holding you back (see Box 1), then develop strategies to overcome your resistance to writing (see Rules 1–10 and Box 2). Also, periodically accelerate your writing practice by joining a writing retreat

### Box 2. Scientists can improve their written productivity using Rules 1–10 and advice from other writers

Tips about writing from writers

"The scariest moment is always just before you start."—Stephen King [10]

"Your desire to write grows with writing."—Desiderius Erasmus [11]

"Write freely and as rapidly as possible and throw the whole thing on paper. Never correct or rewrite until the whole thing is done. Rewrite in process is usually found to be an excuse for not moving on. It also interferes with flow and rhythm which can only come from a kind of unconscious association with the material. . ."—John Steinbeck [12]

"Just write every day of your life. Read intensely. Then see what happens. Most of my friends who are put on that diet have very pleasant careers."—Ray Bradbury [13]

"This is how you do it: You sit down at the keyboard and you put one word after another until it's done. It's that easy, and that hard."—Neil Gaiman [14]

"Almost all good writing begins with terrible first efforts. You need to start somewhere."—Anne Lamott [15]

"My ideas usually come not at my desk writing but in the midst of living."—Anais Nin [16]

"Amateurs sit and wait for inspiration, the rest of us just get up and go to work."—Stephen King [17]

"You can't think yourself out of a writing block; you have to write yourself out of a thinking block."—John Rogers [18]

"There is no failure unless one stops."—Ray Bradbury [19]

or planning a weekend away to write. Truly, the only way to maintain your writing productivity is to keep writing.

## Conclusion

Writing is one of the most important activities that we as scientists engage in because it is critical to share our findings both within and beyond our research community. Use these 10 simple rules to write more and increase your impact as a scientist. But like the iterative process of writing, also evaluate your writing practice often (see Rule 10). Note that it may not be necessary to follow all the rules all the time; experiment and find what works for your writing practice. If you find your writing sessions are becoming less productive because you are facing new hurdles to writing (see Box 1), refine your writing practice using Rules 1–10.

## References


1. Schimel Joshua. Writing Science: How to Write Papers that Get Cited and Proposals that Get Funded. ( Oxford University Press, USA, 2012), 4.
2. Montgomery Scott L. The Chicago Guide to Communicating Science. ( University of Chicago Press, Chicago, IL, 2003), 41.
3. The Eloquent Science Sites, Quotes from Experts on Effective Scientific Writing, 2009. Available from: http://eloquentscience.com/2009/08/quotes-from-experts-on-effective-scientific-writing/. [cited 2018 April 4].
4. SMART Criteria. Wikipedia. Available from: https://en.wikipedia.org/wiki/SMART_criteria. [cited 2017 December 21].
5. Sainani Kristin, Writing in the Sciences. Available from: https://lagunita.stanford.edu/courses/Medicine/SciWrite-SP/SelfPaced/about. [cited 2017 December 21].
6. Zhang W. (2014) Ten Simple Rules for Writing Research Papers. PLoS Comput Biol 10(1): e1003453. https://doi.org/10.1371/journal.pcbi.1003453 https://doi.org/10.1371/journal.pcbi.1003453 PMID: 24499936

7. Mensh B and K. Kording. (2017) Ten Simple Rules for Structuring Papers. PLoS Comput Biol 13(9): e1005619. https://doi.org/10.1371/journal.pcbi.1005619 https://doi.org/10.1371/journal.pcbi.1005619 PMID: 28957311
8. Welch Gilbert H. (1999) Preparing Manuscripts for Submission to Medical Journals: The Paper Trail. Effective Clinical Practice. 2:131–137. PMID: 10538262
9. Strunk, William Jr. The Elements of Style. Pearson Education Limited (England, 2014).
10. King, Stephen. On Writing: A Memoir of the Craft. 10th Anniversary Edition (New York: Pocket Books, 2000), 274.
11. Lilless M Shilling., and Fuller Linda K., editors. Dictionary of Quotations in Communications ( Westport, Connecticut: Greenwood Press, 1997), 276.
12. Quoted in George Plimpton, "John Steinbeck, The Art of Fiction No. 45 (Continued)," The Paris Review, 1975.
13. Perrin, Timothy. (1986) "Ray Bradbury's Nostalgia for the Future," Writer's Digest, February.
14. The Neil Gaiman Journal Sites. Pens, Rules, Finishing Things and Why Stephin Merritt is Not Grouchy, 2004. Available from: journal.neilgaiman.com/2004/05/pens-rules-finishing-things-and-why.asp. [cited 2017 Sept. 9].
15. Lamott Anne. Bird by Bird: Some Instruction on Writing and Life. 1st Edition ( New York: Anchor Books, 1994), 25.
16. Nin, Anais. The Diary of Anais Nin Volume 3 1939:1944. (San Diego, California: A Harvest/HBJ Book, 1969) Vol. 3, 29.
17. King Stephen. On Writing: A Memoir of the Craft. 10th Anniversary Edition ( New York: Pocket Books, 2000).
18. The Kung Fu Monkey Sites. So Where the Hell Have you Been?, 2011. Available from: kfmonkey.blogspot.com/2011/06/so-where-hell-have-you-been.html. [cited 2017 September 9].
19. Bradbury Ray. Zen of the Art of Writing: Releasing the Creative Genius Within You. ( New York: Bantam, 1992), 133.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for responsible referencing

**Bart Penders***

Maastricht University, Care and Public Health Research Institute (CAPHRI), Department of Health, Ethics & Society, Maastricht, the Netherlands

* b.penders@maastrichtuniversity.nl

OPEN ACCESS

**Citation:** Penders B (2018) Ten simple rules for responsible referencing. PLoS Comput Biol 14(4): e1006036. https://doi.org/10.1371/journal.pcbi.1006036

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** April 12, 2018



**Funding:** The work that lead to this publication was, in part, supported by the ZonMW programme Fostering Responsible Research Practices, grant no. 45001005. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

We researchers aim to read and write publications containing high-quality prose, exceptional data, arguments, and conclusions, embedded firmly in existing literature while making abundantly clear what we are adding to it. Through the inclusion of references, we demonstrate the foundation upon which our studies rest as well as how they are different from previous work. That difference can include literature we dispute or disprove, arguments or claims we expand, and new ideas, suggestions, and hypotheses we base upon published work. This leads to the question of how to decide which study or author to cite, and in what way.

Writing manuscripts requires, among so much more, decisions on which previous studies to include and exclude, as well as decisions on how exactly that inclusion takes place. A well-referenced manuscript places the authors' argument in the proper knowledge context and thereby can support its novelty, its value, and its visibility. Citations link one study to others, creating a web of knowledge that carries meaning and allows other researchers to identify work as relevant in general and relevant to them in particular.

On the one hand, citation practices create value by tying together relevant scientific contributions, regardless of whether they are large or small. In the process, they confer or withhold credit, contributing to the relative status of published work in the literature. On the other hand, citation practices exist in the context of current regimes of evaluating science. While it may go unnoticed in daily writing practices, the act of including a single reference in a study is thus subject to value-based criteria internal to science (e.g., content, relevance, credit) and external to science (e.g., accountability, performance).

Accordingly, referencing is not a neutral act. Citations are a form of scientific currency, actively conferring or denying value. Citing certain sources—and especially citing them often—legitimises ideas, solidifies theories, and establishes claims as facts. References also create transparency by allowing others to retrace your steps. Referencing is thus a moral issue, an issue upon which multiple values in science converge. Citing competitors adds to their profiles, citing papers from a specific journal adds to its impact factor, citing supervisors or lab mates helps build your own profile, and citing the right papers helps establish your familiarity with the field. All of these translate into pressures on scientists to cite specific sources, from peers, editors, and others. Fong and Wilhite demonstrate the abundance of so-called coercive citation practices [1]. Also, citation-based metrics have proliferated as proxies for quality and impact over the years [2–4], only to be currently subjected to significant and highly relevant critique [5–8]. To cite well, or to reference responsibly, is thus a matter of concern to all scientists.

Here, I offer 10 simple rules for responsible referencing. Scientists as authors produce references, and as readers and reviewers, they assess and evaluate references. Through this symmetrical relationship to literature that all scientists share, they take responsibility for tying together all knowledge it contains. Producing and evaluating references are, however, distinct processes, warranting different responsibilities. Respecting this dual relationship researchers have with literature, the first six rules primarily refer to producing a citation and the responsibilities this entails. The second set of four rules refers to evaluating citations and the meaning they have or acquire once they have become part of a text.

## Rule 1: Include relevant citations

All scholarly writing requires a demonstration of the relevance of the questions asked, a display of the methods used, a rationale for the use of materials, and a discussion of issues relevant to the content of the publication. All of these are done, at least in large part, by including citations to relevant previous work. Omitting such references can wrongfully suggest that your own publication is the origin of an idea, a question, a method, or a critique, thereby illegitimately appropriating them. Citations identify where ideas have come from, and consulting the cited works allows readers of your text to study them more closely, as well as to evaluate whether your use of them is appropriate.

A single exception exists when facts, findings, or methods have become part of scientific or scholarly canon. There is no need to include a citation on the claim that DNA is built out of four bases, nor do you have to cite Kjell Kleppe or Kary Mullis every time you use PCR (neither do I right now). However, the decision as to when something truly becomes part of canon can be quite difficult and will include periods of adjustment (with irregular citation) and negotiation (on whether to cite or not).

## Rule 2: Read the publications you cite

Citation is not an administrative task. First, a single paper can be cited for multiple reasons, ranging from reported data to methods, and can be cited both positively and negatively in the literature. The only way to identify whether its content is relevant as support for your claim is to read it in full.

Second, the collection of citations included to support your work and argument is one of the elements from which your work draws credibility. The same goes for the citations you include to criticise, dispute, or disprove. As a consequence, a chain is only as strong as its weakest link. The quality of the publication you trust and upon which you confer authority codetermines the quality and credibility of your work. Citation rates, especially on the journal level, do not correspond well to research quality [9], and they conflate positive and negative citations, not distinguishing authority conferred or authority that is challenged. To cite meaningfully and credibly requires that you consult the content of a publication rather than whether others have cited it, as a criterion for citation.

## Rule 3: Cite in accordance with content

If, at some phase in the research, you have decided that a specific study merits citation, the issue of specifically how and where to cite it deserves explicit consideration. Mere inclusion does not suffice. Sources deserve credit for the exact contribution they offer, not their contribution in general. This may mean that you need to cite a single source multiple times throughout your own argument, including explanations or indications why.

A specific way to break Rule 3 is in the form of the so-called 'Trojan citation' [10]. The Trojan citation arises when a publication reporting similar findings to your own is cited in the context of a discussion of a minor issue, ignoring (sometimes deliberately) its key argument or contribution. By focussing on a trivial detail, the Trojan citation obscures the true significance of the cited work. As a consequence, it hides that your work is not as novel as it seems. As a questionable citation practice, a Trojan citation can be used to satisfy reviewers' or editors' requests to include a reference to a relevant paper. Alternatively, a Trojan citation may emerge unknowingly when (1) you are unaware of the content of a cited publication (not adhering to Rule 2 creates a very significant risk of being unable to follow Rule 3) or (2) disputes exist in the scientific community or among the authors on the contribution and/or quality of a scientific publication (in which case, Rule 4 will help).

## Rule 4: Cite transparently, not neutrally

Citing, even in accordance with content, requires context. This is especially important when it happens as part of the article's argument. Not all citations are a part of an article's argument. Citations to data, resources, materials, and established methods require less, if any, context. As part of the argument, however, the mere inclusion of a citation, even when in the right spot, does not convey the value of the reference and, accordingly, the rationale for including it. In a recent editorial, the *Nature Genetics* editors argued against so-called neutral citation. This citation practice, they argue, appears neutral or procedural yet lacks required displays of context of the cited source or rationale for including [11]. Rather, citations should mention assessments of value, worth, relevance, or significance in the context of whether findings support or oppose reported data or conclusions.

This flows from the realisation that citations are political, even though that term is rarely used in this context. Researchers can use them to accurately represent, inflate, or deflate contributions, based on (1) whether they are included and (2) whether their contributions are qualified. Context or rationale can be qualified by using the right verbs. The contribution of a specific reference can be inflated or deflated through the absence of or use of the wrong qualifying term ('the authors suggest' versus 'the authors establish'; 'this excellent study shows' versus 'this pilot study shows'). If intentional, it is a form of deception, rewriting the content of scientific canon. If unintentional, it is the result of sloppy writing. Ask yourself why you are citing prior work and which value you are attributing to it, and whether the answers to these questions are accessible to your readers.

## Rule 5: Cite yourself when required

In the context of critical discussions of citations and evaluations of citation-based metrics, self-citation has almost become a taboo. It is important to realise, though, that self-citation serves an important function by showing incremental iterative advancement of your work [12]. As a consequence, your previous work or that of the group in which you are embedded should be cited in accordance with all of the rules above. The amount of acceptable self-citation is very likely to differ between fields; smaller fields (niche fields) are likely to (legitimately) exhibit more.

This does not mean that self-citation is always unproblematic. For instance, excessive self-citation can suggest salami slicing, a publication strategy in which elements of a single study are published separately [13]. This questionable research practice, in tandem with self-citation, aims to inflate publication and citation metrics.

## Rule 6: Prioritise the citations you include

Many journals have restrictions on the number of references authors are allowed to include. The exact number varies per publisher, journal, and article type and can be as low as three (for a correspondence item in *Nature*). Even if no reference limit exists, other journals impose a word limit that includes references, effectively also capping the amount of references. Coping with these limits sometimes requires difficult decisions to omit citations you may feel are legitimate or even necessary. In order to deal with this issue and avoid random removal of

### Box 1: Reference prioritisation

'Ten simple sub-rules for prioritising references' can help to facilitate prioritisation. In most cases, a subset of the 10 sub-rules will suffice. First, prioritise anew for each publication. Prioritisations cannot (easily) be copied from one study to another. Second, prioritise per section (e.g., introduction, methods, discussion), not across the entire paper. Different sections require different types of support. Third, for the introduction, prioritise reviews, allowing broad context for relevance and aim. Fourth, for the discussion, prioritise empirical papers, allowing detailed accounts of relative contribution. Fifth, prioritise reviewed over un- or prereviewed papers (e.g., editorials, preprints, etc.). Sixth, deprioritise self-citations. Seventh, limit the number of citations to support a specific claim, if necessary, to a single citation. Eighth, move methodological citations to supplementary (online) information. Ninth, in cases of equal relevance, prioritise citation of female first or last authors to help repair gender imbalances in science. Tenth, request the inclusion of additional references with the editors, arguing that you have used all of the previous nine sub-rules.

references, all desired citations require prioritisation. A few rules of thumb, shown in Box 1, will help decisions on reference priority.

## Rule 7: Evaluate citations as the choices that they are

Research publications are not mere vessels of data or findings. They convey a narrative explaining why questions are worth asking, what their answers may mean, how these answers were reached, why they are to be trusted, and more. They also have a purpose in the sense that they will act as support for other studies to come. Each of the elements of their story is supported by links to other studies, and each of those links is the result of an active choice by the author(s) in the context of the goal they wish to achieve by their inclusion.

At the other end of the narrative, readers assess and evaluate the story constantly, asking whether it could have been told differently. The realisation that narratives can be told differently, supported by other citations to other prior work, does not disqualify them. Both the story and the choice of citations are political choices meant to provide the argument with as much power, credibility, and legitimacy the author(s) can muster. They are tailored to the audience the authors seek to convince: their peers. The choice to include or exclude a reference can only be evaluated in the context of that narrative and the role they play in it. Peritz has provided a classification of citation roles to assist this evaluation [14].

## Rule 8: Evaluate citations in their rhetorical context

Rhetorical strategies serve to convince and persuade. Narratives are but one of the tools that can be used to persuade audiences. Metaphors, numbers, and associations all feature in our research papers as tools to convince our readers. The genre of the scientific article has had centuries to evolve to incorporate many of them, with the goal of convincing readers that the author is right. Bazerman has literally written the book on this [15] and urges us to consider academic texts and their features as part of social and intellectual endeavours. Citations are a part of the social fabric of science in the sense that through citing specific sources, authors show their allegiance to schools of thought, communities, or, in the context of scientific controversies, which paradigm they consider themselves part of. Other rhetorical uses of citations include explicit citations to notable figures and their work, which can serve as appeals to authority, while long lists of citations can serve as proxies for well-studied subjects.

Consider the following: Authors can describe a field as well-studied and include three references—X, Y, and Z—as support for their claim. Alternatively, they can argue that a field is understudied but that three exceptions exist, i.e., X, Y, and Z. Understanding the value attributed to X, Y, and Z in that particular text requires assessment of the rhetorical strategies of the author(s).

## Rule 9: Evaluate citations as framed communication

Authors use words to accomplish things and, in service of those goals, position their work and that of others. They frame prior work in a very specific way, supporting the arguments made. We all do. The positioning of X, Y, and Z either as the norm or as exceptions, as shown in Rule 8, is an example of framing. It is important to recognise such framing and that X, Y, and Z acquire meaning in the text as the result of the frame. There is no frameless communication, as Goffman [16] demonstrated. All messages and texts contain and require a frame—a structure of definitions and assumptions that help organise coherence, connections, and, ultimately, meaning—or in other words, a perspective on reality.

As a result, a citation is not a neutral line drawn between publications A and B. Rather, the representation of cited article A only acquires meaning in the context of citing in article B. Article A can be framed differently when cited in work B or C. It can be framed as innovative in B or dogmatic in C. Framing usually is not lying or deceiving; it is a normative positioning of evidence in context. Hence, a citation is a careful translation of a source's relevant elements, which acquire meaning in that context only.

An important consequence of this is that merely counting citations of article A in the literature does not inform us of the value (or many types of value or lack thereof) of article A to the scientific community. This point also appears as the first principle in the Leiden Manifesto, which argues that quantitative metrics can only support qualitative metrics (i.e., reading with an attentive eye for politics, rhetoric, context, and frame—or as adhering to Rules 7–9). The Leiden Manifesto was published by bibliometricians and scholars of research evaluation following the 2014 conference on Science and Technology Indicators in Leiden, the Netherlands. It warns against the abuse of, among other things, citation-based research metrics [9].

## Rule 10: Accept that citation cultures differ across boundaries

Despite critiques of the system, science is organised in such a way that citations continue to act as a currency that is represented as being universal [4]. However, citation practices are, for the most part, local practices, whether local to laboratories or department or local to disciplines. The average number of citations per paper differs between disciplines, and the way that citations are represented in the text and the value of being cited also differ radically [17]. What counts as proper citation practice in molecular biology—for instance, the inclusion of multiple references following a statement—is considered unacceptable in research ethics or science policy, in which single references require paragraphs of contextualisation and translation (see Rule 9). When reading a paper from an adjacent discipline, respect its different norms and conventions for responsible referencing and proper citation. If you are cited by a scientist from another discipline, assess that act as existing in a (however slightly) different citation culture.

## Acknowledgments

I thank Maurice Zeegers and his team, who work on citation analyses, for stimulating me to think about the issue of citation more clearly, deeply, and critically, resulting in the considerations above. I also thank David Shaw for critical comments, moral support, and editorial assistance. As a closing note, as the human being that I am, I too have quite possibly referenced imperfectly in my previous work.

## References

1. Fong E, Wilhite A. Authorship and citation manipulation in academic research. PloS ONE. 2017; 12(12): e0187394. https://doi.org/10.1371/journal.pone.0187394 PMID: 29211744
2. Feller I. Performance measurement and the governance of American academic science. Minerva. 2009; 47(3):323.
3. Garfield E, Merton R. Citation indexing: Its theory and application in science, technology, and humanities: New York: Wiley; 1979.
4. Wouters P. The citation culture. Amsterdam: University of Amsterdam; 1999.
5. Dahler-Larsen P. Constitutive effects of performance indicator systems. Dilemmas of engagement: Evaluation and the new public management. Emerald Group Publishing Limited; 2007. p. 17–35.
6. De Rijcke S, Wouters P, Rushforth A, Franssen T, Hammarfelt B. Evaluation practices and effects of indicator use—a literature review. Research Evaluation. 2016; 25(2):161–9.
7. Bornmann L, Daniel H-D. What do citation counts measure? A review of studies on citing behavior. Journal of documentation. 2008; 64(1):45–80.
8. Müller R, de Rijcke S. Exploring the epistemic impacts of academic performance indicators in the life sciences. Research Evaluation. 2017; 26(3):157–68.
9. Hicks D, Wouters P, Waltman L, De Rijcke S, Rafols I. Bibliometrics: the Leiden Manifesto for research metrics. Nature. 2015; 520:429–31. https://doi.org/10.1038/520429a PMID: 25903611
10. Shaw D. The Trojan Citation and the "Accidental" Plagiarist. Journal of Bioethical Inquiry. 2016; 13 (1):7–9. https://doi.org/10.1007/s11673-015-9696-7 PMID: 26780105
11. Nature Genetics. Neutral citation is poor scholarship. Nature Genetics. 2017; 49:1559. https://doi.org/10.1038/ng.3989 PMID: 29074946
12. Hyland K. Self-citation and self-reference: Credibility and promotion in academic publication. Journal of the Association for Information Science and Technology. 2003; 54(3):251–9.
13. Jackson D, Walter G, Daly J, Cleary M. Multiple outputs from single studies: Acceptable division of findings vs. 'salami' slicing. Journal of clinical nursing. 2014; 23(1–2):1–2. https://doi.org/10.1111/jocn.12439 PMID: 24313937
14. Peritz B. A classification of citation roles for the social sciences and related fields. Scientometrics. 1983; 5(5):303–12.
15. Bazerman C. Shaping written knowledge: The genre and activity of the experimental article in science. Madison, WI: University of Wisconsin Press; 1988.
16. Goffman E. Frame analysis: An essay on the organization of experience. Cambrdige, MA: Harvard University Press; 1974.
17. Aksnes D, Rip A. Researchers' perceptions of citations. Research Policy. 2009; 38(6):895–905.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for writing a popular science book

**Adam J. Kucharski***

Centre for the Mathematical Modelling of Infectious Disease, London School of Hygiene and Tropical Medicine, London, United Kingdom

* adam.kucharski@lshtm.ac.uk

OPEN ACCESS

**Citation:** Kucharski AJ (2018) Ten simple rules for writing a popular science book. PLoS Comput Biol 14(2): e1005808. https://doi.org/10.1371/journal.pcbi.1005808

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** February 1, 2018



**Funding:** The author received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

Scientists have a long history of writing for a general audience [1]. By communicating ideas and discoveries to the wider public, popular science books can generate debate, influence culture, and inspire future researchers. They can also provide new skills and opportunities, and in some cases a second career, for their authors [2].

Writing scientific papers is a central part of academic life, but there are some notable differences when it comes to publishing books for a wider audience. Based on my experience of writing and contributing to popular science books [3–5], the following rules summarise the main points a budding author needs to consider, from initial proposal through to postrelease publicity.

## Rule 1: Build writing experience

There's an old adage that to write a good novel, you should write a complete draft, put it in a drawer, then forget about it and start work on what will become the real book. Nonfiction is similar. Writing takes a huge amount of practice, which means it's a good idea to first build up a portfolio of articles, stories, and blog posts. Not all of these attempts will work out; it's likely that many of your words will go unpublished. However, writing regularly means you can explore and refine your style. In the process, you'll develop many of the skills necessary for book writing, from spotting stories and pitching ideas to structuring and editing content. It will also help you decide whether you really want to take on a book project.

## Rule 2: Find the right topic

If you have an idea for a book, there are two main questions you should consider. Why does this book need to be published now? And why are you the person to write it? It helps if the topic is timely and new (or presents a new take on a familiar issue). Try thinking about what's important in the medium term: your eventual publication date may be three to four years away, so what's currently in the news may not still be topical. You also need to think about how your academic background relates to the book. Although it might seem easiest to focus on your specific area of research, there are benefits to telling a wider story. First, it's more likely to be relevant to a general reader, who may be broadly interested in your field but not so keen on a whole book about one subtopic. Second, it can make book writing more interesting for you because you'll be able to discover new things as you research and write.

## Rule 3: Get an agent

If you want to write a trade book (i.e., for a general audience), it helps to have a literary agent. Although they typically take around a 15% cut of your income, most publishers won't look at a submission unless it comes from a reputable agent, and it's better to have 85% of a book advance from a good publisher than 100% of nothing. Agents will also edit the book proposal, pitch to publishers, negotiate contracts, advise on potential problems, and handle royalties.

How do you get an agent? It helps if you have a strong book idea and you've recently done something that stands out. If you're an established academic, you may have high-profile discoveries and projects to your name; if you're more junior, you might have won an award or given talks that have gained a lot of attention. Remember, agents and publishers are looking to sign you as well as the book.

## Rule 4: Write a proposal

Nonfiction books are generally submitted to publishers in the form of a proposal. Like research grants, there are certain conventions that must be followed [6]. Proposals usually consist of about 5,000 words outlining what the book will cover, why you are qualified to write it, and what the competition is. Authors will often submit a sample chapter as well, usually from midway through the book, to give the publisher an idea of their writing style (which means another few thousand words). If this sounds like a lot, spare a thought for fiction writers, who generally have to finish the whole novel first. A good agent will help you edit the proposal to get it in as good a shape as possible. Although you should listen to their advice, it's important that you're comfortable with the focus of the book. Don't add things that you aren't happy (or qualified) to write about.

## Rule 5: Pitch to publishers

Once the proposal is ready, your agent will send it to dozens of editors at different publishers. This will likely lead to dozens of rejection e-mails. Despite the classic stories of vicious rejections (Knopf famously called Orwell's *Animal Farm* 'stupid and pointless' [7]), rejections are often polite and reasonable. Some editors may have recently acquired a book on a similar topic or had a bad run with that genre in the past. Or they may like the idea, but their sales team—who will eventually be pitching the book to retailers—do not.

If an editor is interested, they'll often get you on the phone to find out more about your plans and give you the chance to find out more about theirs. Then, if all is well, the offers will start coming in. Publishers typically offer an advance against royalties, with the money spread over several payments, e.g., payment on signature, manuscript delivery, hardback, and paperback publication. So, a big headline number may ultimately be split over three or more years. On the plus side, however, there is often potential to sign foreign language deals as well as English ones.

## Rule 6: Find your structure and voice

First person or third person? Narrative driven or explanatory? Lighthearted or serious? Popular science books allow for a much wider range of writing styles than academic papers. In the proposal, you'll have outlined the overall book structure, but you'll also need to decide how individual sections and chapters fit together. You might want to follow certain characters or include anecdotes and history to motivate the scientific content. If you can show the conflicts and struggles that shaped the science, it will help draw your readers into the research [8].

Your writing will also have a personality—your 'voice' on the page—and this will influence the feel of the book. Some writers have very distinctive, and possibly even distracting, voices. In popular science writing, it's particularly important to avoid condenscending to the reader. Although they might not know scientific jargon, they should be able to grasp the concepts if you describe them well. As the old media adage goes, never overestimate the knowledge of your readers or underestimate their intelligence [9].

## Rule 7: Research and interview

A typical 250-page popular science book will contain around 75,000 words. If you agree to deliver the first draft in 18 months, that works out at about 1,000 words a week on average. However, behind those words lies a mountain of research and interviews. Depending on the focus of the book, you might need to sift through archives, newspapers, and biographies or even run simulations and analyse data. When researching a topic, it's worth tracking down the primary source when possible. Just as facts in academic papers can wander astray over the course of several citations [10], well-known quotes and stories may turn out to be apocryphal [11].

It helps to interview people who are involved in the subject area of your book. Your job may work in you favour here: researchers can be happier to talk to a fellow academic than a journalist. However, it's still important to follow good journalistic practice. Interviews should be on the record, and you should record your conversations to ensure accurate quotes [12]. Researching, interviewing, and writing take a lot of effort, which must be balanced against academic work. Good time management is essential: set aside specific evenings, weekends, and holidays for the book. In my experience, a good writing day will add around 500 to 1,000 words.

## Rule 8: Edit and edit again

Once the initial draft is ready, it will need several rounds of editing. Many of the rules for writing and editing papers [13,14] also apply to popular science books. Avoid long, complex sentences and adverbs. Make sure there is a clear logical structure. Get feedback from colleagues who are not afraid to critique the work. You'll probably spend most time researching and writing about the topics you don't know so well, so pay particular attention to sections of the book where you're on familiar ground; this is where simple errors can creep in. Make sure you also understand what constitutes copyright infringment and libel, especially if you're covering controversial events [15]. Your publisher should also be able to help advise on this.

When you send the full manuscript to your publisher, there will be two main editing steps. As with any publication, the main editor will advise on broad aspects like content and structure; copyeditors later deal with the grammar and wording. If you've signed both a United States and United Kingdom deal, you may also need to 'translate' the manuscript for audiences either side of the Atlantic.

## Rule 9: Plan your publicity

Around six months before the publication date, your publisher will start ramping up for the launch. They'll usually send dozens of review copies of the book to a range of reporters, producers, and editors. They'll also mail copies to well-known authors and academics to collect 'blurbs' (those complimentary quotes you often see on first editions of books). Most publishers have in-house publicists, and they'll help you pitch book-related articles to newspapers and magazines [16]. These are typically around 700 to 800 words, highlighting stories and ideas from the book that would be of interest to a particular readership.

Very occasionally, a debut book will sell lots of copies without much publicity. However, there seems to be a pretty strong link between sales and traditional publicity efforts. If a book is reviewed or featured in a major newspaper or mentioned on radio or TV, it can shoot up the Amazon rankings.

## Rule 10: Do lots of speaking

Near the book launch date, your publisher can help line up interviews, which may include radio, podcasts, newspapers, and television. If you already have contacts in the media, it's a good idea to get in touch and see if they'd be interested, too. There may also be the option to give public talks about the book. These might be one-off events or technology- and/or business-focused conferences with several other speakers. Each event will involve slightly different time slots and audiences, so although you may want to include a few recurring topics, you should plan to be flexible.

It can be difficult to avoid publicity conflicting with your scientific work. In some cases, you may need to turn down book-related events because of existing academic commitments. This promotion stage is a lot of work, but it can also be a lot of fun. After months of working alone on the book, it's an opportunity to share the stories and ideas in your book with a wider audience. As an author, it's a wonderful feeling to see people enjoy something you've created.

## References


1. Culture Lab. What popular science books have changed the world? New Scientist 2012. Available from: https://www.newscientist.com/blogs/culturelab/2012/08/what-popular-science-books-changed-the-world.html. Accessed on 10 August 2017.
2. Kreeger KY. Writing Science. The Scientist 2000. Available from: http://www.the-scientist.com/?articles.view/articleNo/12626/title/Writing-Science/. Accessed on 10 August 2017.
3. Kucharski AJ. The Perfect Bet: How Science and Math Are Taking the Luck Out of Gambling. 1st ed. Basic Books 2016
4. Kucharski AJ. The future of medicine. In Al-Khalili J. (Ed.), What's Next? Profile Books 2017
5. Kucharski AJ. Finding Apollo. In Parc S. (Ed.), 50 Visions of Mathematics. Oxford University Press 2014
6. Bourne PE, Chalupa LM. Ten Simple Rules for Getting Grants. PLoS Comput Biol 2006; 2:e12 https://doi.org/10.1371/journal.pcbi.0020012 PMID: 16501664
7. Watercutter A. The Rejection Letters That Burned Everyone From George Orwell to Aspiring Clowns. Wired 2014. Available from: https://www.wired.com/2014/01/sundance-free-fail-instagram/. Accessed on 10 August 2017.
8. Martinez-Condea S, Macknik SL. Finding the plot in science storytelling in hopes of enhancing science communication. Proc Natl Acad Sci USA. 2017; 114:8127–8129 https://doi.org/10.1073/pnas.1711790114 PMID: 28765506
9. Radford T. A manifesto for the simple scribe–my 25 commandments for journalists. Guardian Online 2011. Available from: https://www.theguardian.com/science/blog/2011/jan/19/manifesto-simple-scribe-commandments-journalists. Accessed on 10 August 2017.
10. Reich NG, Perl TM, Cummings DAT, Lessler J. Visualizing Clinical Evidence: Citation Networks for the Incubation Periods of Respiratory Viral Infections. PLoS ONE 2011; 6:e19496 https://doi.org/10.1371/journal.pone.0019496 PMID: 21559339
11. O'Toole G. Hemingway Didn't Say That: The Truth Behind Familiar Quotations. 1st ed. Little A 2017
12. Reuters. The Essentials of Reuters sourcing. Available from: http://handbook.reuters.com/?title=The_Essentials_of_Reuters_sourcing. Accessed on 10 August 2017.
13. Zhang W. Ten Simple Rules for Writing Research Papers. PLoS Comput Biol 2014; 10:e1003453 https://doi.org/10.1371/journal.pcbi.1003453 PMID: 24499936
14. Weinberger CJ, Evans JA, Allesina S. Ten Simple (Empirical) Rules for Writing Science. PLoS Comput Biol 2015; 11:e1004205 https://doi.org/10.1371/journal.pcbi.1004205 PMID: 25928031
15. Bingham H. The Writers' and Artists' Yearbook Guide to Getting Published: The Essential Guide for Authors. Bloomsbury 2010
16. Pitch Database. The Open Notebook 2017. Available from: http://www.theopennotebook.com/pitches/. Accessed on 10 August 2017.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for drawing scientific comics

**Jason E. McDermott[1,2]*, Matthew Partridge[3], Yana Bromberg[4,5]**

**1** Computational Biology and Bioinformatics, Pacific Northwest National Laboratory, Richland, Washington, United States of America, **2** Department of Molecular Microbiology and Immunology, Oregon Health & Sciences University, Portland, Oregon, United States of America, **3** Engineering Photonics, Cranfield University, Cranfield, Bedfordshire, United Kingdom, **4** Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, New Jersey, United States of America, **5** Institute for Advanced Study, Technische Universität München, Garching, Germany

* jason.mcdermott@pnnl.gov



OPEN ACCESS

**Citation:** McDermott JE, Partridge M, Bromberg Y (2018) Ten simple rules for drawing scientific comics. PLoS Comput Biol 14(1): e1005845. https://doi.org/10.1371/journal.pcbi.1005845

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** January 4, 2018



**Funding:** NSF CAREER Award 1553289; the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

Institutions around the world are in a constant struggle to improve science communication. From calls for journal papers to be simpler and more accessible to encouraging scientists to take a more active role through community engagement, there is a drive to demystify and improve public understanding of and engagement with science [1–3]. This drive for engagement is crucial to both helping recruit the next generation of scientist and highlighting the impact and role science has in public life. It also has a role in peer-to-peer communication and wider dissemination of ideas throughout the community. Technology has greatly helped expand the range of teaching styles that a lecturer can call on to reach more people in new ways. Social media outlets like Twitter, Facebook, Instagram, and Tumblr have expanded the reach of science communication within and across scientific disciplines and to the lay public [1, 3]. These new communication channels seem to support endless innovations in the development of videos, interactive quizzes, and instant feedback. Yet they are also providing a platform for a renaissance of one of the simplest and most effective methods for communicating ideas—comics. There are few scientists who haven't heard of Randall Munroe, the artist behind the web comic "xkcd" [4], which features amazing graphic explanations on everything from climate change [5] to data storage [6]. These comics are widely appealing to a diverse audience and are posted on walls in laboratories and pubs alike. The ideas that they explain are complicated, but by simplifying them down to the core messages and by providing simple visual analogies, the comics educate and engage the groups that other media cannot always reach.

A comic is generally an illustration that employs metaphor and/or storytelling to clearly communicate an idea to a broad audience. Comics often employ humor, but their narratives can be exclusively informational in nature or can deal with nonhumorous topics. Comics can take multiple forms, from the single panel one-liner, to multiple panels, to graphic novels that span multiple pages. There are a number of science- and academic-oriented comics in circulation, including xkcd, PHD [7], and the authors' own Errant Science [8] and RedPen/BlackPen [9].

An effective comic can communicate difficult ideas efficiently, illuminate obscure concepts, and create a metaphor that can be much more memorable than a straightforward description of the concept itself. Comics can be used to punctuate presentations or journal publications [10–12] to increase impact. In public health education, comics have long been recognized as an effective tool for reaching lots of different populations for education on subjects like cancer [13], fitness [14], and diabetes [15], to name only a few. A recent trend is for scientists and artists (and scientist-artists) to capture the content of talks at conferences, or indeed entire meetings [16], as graphical notes [17]. A vibrant and growing scientific community on social media makes this a particularly effective method for expanding the intended audience; i.e., particularly engaging comics are "virally" spread within very short time frames. Science comics have also been included in research studies to enhance the story and facilitate understanding by a broader audience [10–12]. Certain journals have a "cartoon" category for submission so that the comic will appear in a citable form in publication [18]. Broadly, all of these avenues represent different ways of promoting work to others.

Here, we focus on three key opportunities provided by comics. First, presenting ideas visually is an effective entry point to complex ideas. Second, using metaphor makes information memorable in ways that literal descriptions do not. Third, though not all topics and situations are suited to the use of humor, employing humor can engage nonexperts and experts alike. It both reduces the levels of intimidation associated with presenting scientific results to a wide audience and breaks down the barriers to understanding that often come with new science.

Here, we set out several guidelines that we hope will convince more scientists that drawing your own comics is simpler than you think. We start with breaking the biggest deterrent of all.

## Rule 1: You don't have to be good at art

Comics are not about art. They are about conveying a message in graphic form. Graphs and plots are for accurately conveying data, diagrams are for accurately depicting a system or setup, and comics are there to help people understand an idea. Some of the best cartoonists and comic artists cannot draw much better than wobbly lines forming strange shapes (Figs 1–10). The trick is to find the shapes that best convey the point you are trying to make. For example, you can convey the sense of scale within a system with a single circle and a dot. Use the dot to represent your smallest scale and then draw a proportionally scaled circle to represent the larger scale. This very basic comic conveys a sense of scale better than writing "small" and "twenty times bigger" (Fig 1). As is explored further, it's not about the smoothness of the lines or the accuracy of the circles, and if you can make a crude shape on paper, you can do what we set out in these rules. Anyone can create a comic, and often the biggest barrier is just getting over the idea that you can't. With practice, you'll get better at communicating ideas this way.

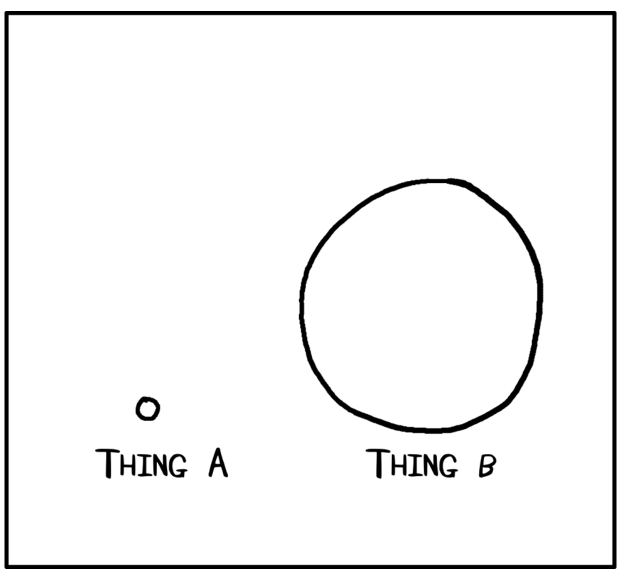



**Fig 1. Sense of relative scale can be conveyed with very simple drawings.**

https://doi.org/10.1371/journal.pcbi.1005845.g001


Fig 2. Adding information can create a "Vennster" (the intersection of a Venn diagram and monster).

https://doi.org/10.1371/journal.pcbi.1005845.g002

While a piece of paper and a pencil are enough to get started drawing, there are also numerous websites that provide comic drawing software free [19] as well as guides on some of the finer details behind producing full comics [20].

## Rule 2: Comics should be simple

The use of comics should make a complicated idea simpler and easier to understand—not harder! Figure out which of your components or steps can be removed or combined in your comic. Comics are like figures in papers; they are best when each conveys one message. Complicated multithreaded comics can look like a "ridiculogram"—a graph with six axes or a Venn

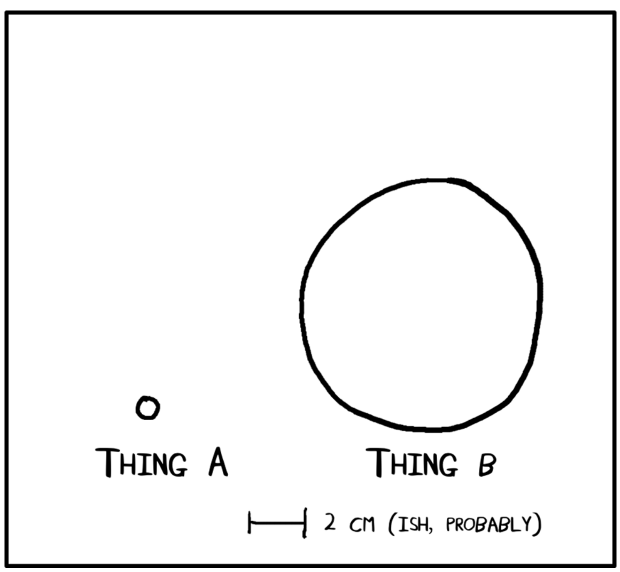



Fig 3. In general use, only enough information to get the idea across.

https://doi.org/10.1371/journal.pcbi.1005845.g003

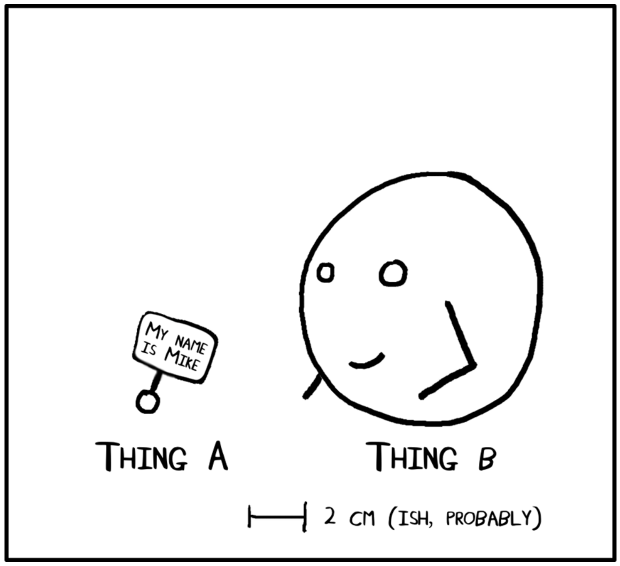



**Fig 4. Adding faces and names increases engagement.**

https://doi.org/10.1371/journal.pcbi.1005845.g004

diagram with six categories, one of them shaped like a banana (see Fig 4 from [21]). These are graphical strategies that are fun to look at but cannot be easily interpreted (Fig 2). As with the previous example, the comic works best when conveying a simple message, in that case indicating the scale of the system.

## Rule 3: Make it right, not perfect

Check the science. If your comic has scientific ideas in it, take the time to make sure you have the details right. If it's mainly just a funny-joke comic, it doesn't need to be absolutely right.

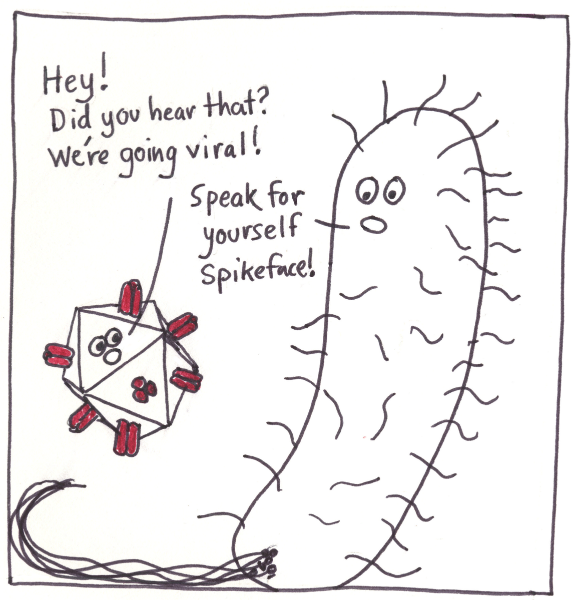



**Fig 5. Comics have a way of going viral.**

https://doi.org/10.1371/journal.pcbi.1005845.g005

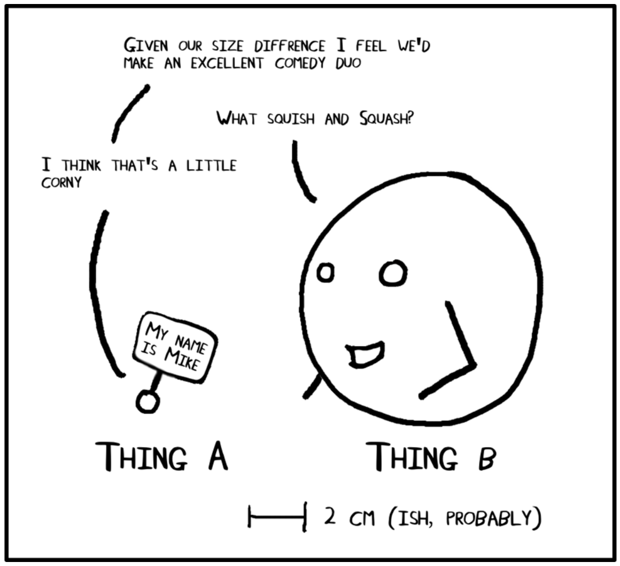



**Fig 6. Interaction between characters is an essential part of storytelling.**

https://doi.org/10.1371/journal.pcbi.1005845.g006

For example, you can add footnotes to comics to point out scientific inaccuracies. But if it's a comic that is meant to really illustrate a scientific concept for the purpose of education, then it should be as factually correct as you can make it. Including incorrect information in something that is intended to educate is misleading and can lead to misconceptions for those you are trying to reach who may not have a scientific background. In the example of the dot and the circle, no one is going to run a volume analysis on your comic (Fig 3). But they will expect it to be within a by-eye–visible order of magnitude of what you are trying to convey.

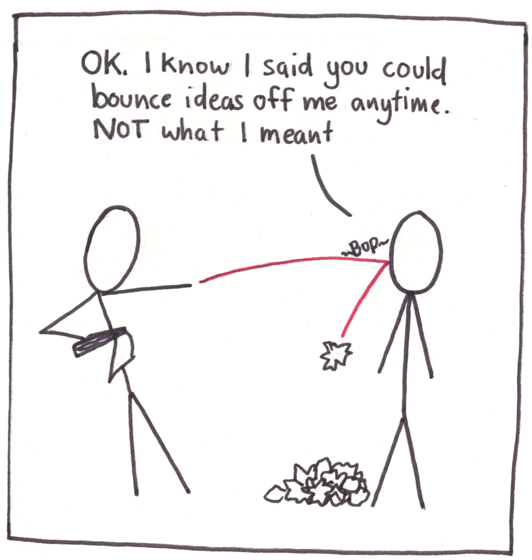



**Fig 7. Find a trusted friend to bounce ideas off of.**

https://doi.org/10.1371/journal.pcbi.1005845.g007

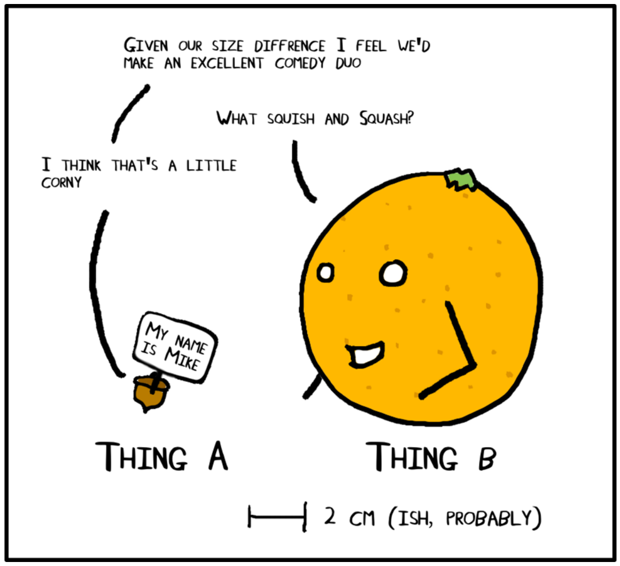



**Fig 8. Adding an analogy can transform a comic.**

https://doi.org/10.1371/journal.pcbi.1005845.g008

## Rule 4: Characters can improve engagement

Create characters with personality that can guide the reader—what your character wears, how tall they are, what they are carrying. If your subjects are inanimate objects, then add personality by including a face. Humans see a face and easily recognize humanity in objects. The famous example is when you hold a pencil, tell everyone that you have named it Steve and then immediately break it [22]. People will tend to feel empathy for the pencil. Simply naming your shapes can be enough to help people engage with the comic and understand and remember the message it conveys. Personification allows the expression of emotions and interactions between players in your comic that let a story be told (see Rule 6). In the dot and circle example, this can be as simple as giving one of the objects hand-like shapes (Fig 4). Or in a more real-world setting, adding something as simple as googly eyes to equipment can produce the same result.

## Rule 5: Don't punch down

Comics have a way of going viral (Fig 5), and it's a good idea to reflect on the possible consequences of everyone in the world reading your comic. (No, not literally everyone in the world.) Don't punch down: making mean fun of those less powerful or privileged than you is bad form, and you should evaluate with every comic you produce. Maintaining a spirit of fun, self-effacing humor and/or commiseration can often express similar ideas without putting anyone down. Be careful with work-inspired comic ideas. Complaining about your workplace using specific details is simply not a good idea. If you do, try not to make any situation or anyone in the comic identifiable—unless you've asked them first or they're a public figure. It shouldn't need to be said, but avoid jokes that are sexist, racist, ableist, or most other "ists." (Marxist jokes may be back on the table.) You should really avoid those in real life as well. If you do get criticized for a comic you've posted, take a deep breath, let it out, find a trusted and honest friend or colleague, and ask their opinion. Don't be afraid to pull the comic. There are rare cases in which any communication, especially those involving social media, has grown to

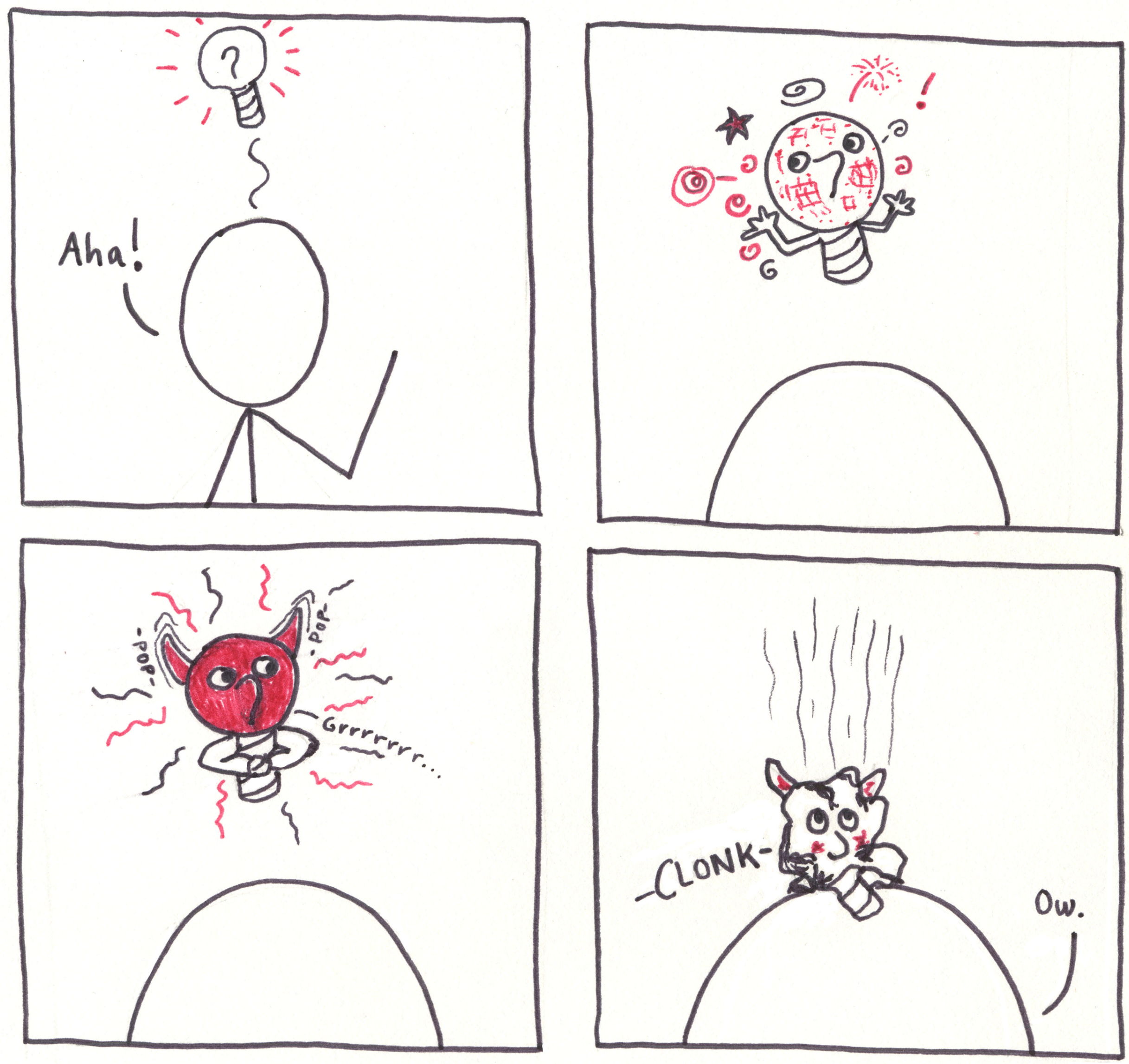



**Fig 9. Some ideas take time to develop, others are better fresh.**

https://doi.org/10.1371/journal.pcbi.1005845.g009

have serious implications for the author [23] and, potentially, the institution they are associated with.

## Rule 6: Tell a story

A good comic, like a good scientific manuscript, tells a story. Like a story, a comic has a beginning (the setup), a middle (the conflict), and a resolution (the punchline). A single-panel comic compresses all these into a single illustration, but it may lay out all the elements of the story in the panel (Fig 6). If illustrating a process or mechanism, start with Rule 4 and personify the elements. Then, think about the story your comic is telling—the steps of the process—

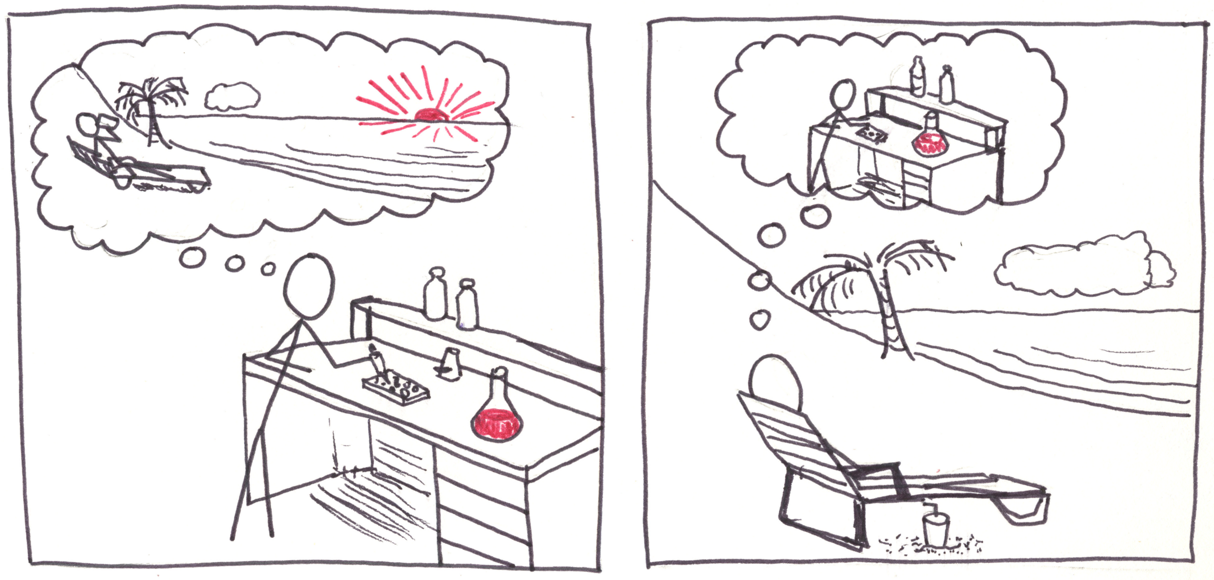


**Fig 10. Relax and have fun—in whatever way you can.**

https://doi.org/10.1371/journal.pcbi.1005845.g010

and how this might be made more memorable by using your characters. What would the enzyme in your comic say if it could talk? You've just given the enzyme that ability! All stories have conflict. This can be in the form of an actual villain, a conflict of ideas, an unseen context to the story, or a joke that the reader is likely to understand. It is important that the language you use to help tell this story be simple and legible. Ideally, it should be tested on nonnative speakers. The impact of the comic can be highly reduced if readers don't understand the dialogue.

## Rule 7: Draw on what you know and find your own voice

As with many other things, the adage "write what you know" applies to comics as well, but don't feel limited to only what you're an expert in. Draw from your own experience (paying attention to Rule 5, of course), and if you are comfortable taking on difficult problems or ideas, then go ahead. Personal stories that come from your own experience and emotions can be incredibly powerful [24]. Your comics might be topical, but that's ok—science is topical. And by bringing something that you care about and understand to a wider audience, you might just communicate outside your subspecialty. Paying attention to concepts you find important, issues that are relevant to you, and interactions you have daily can be a treasure trove of ideas if you pay attention. If you have a comic or an idea for a comic, try bouncing it off a trusted friend or colleague (Fig 7); then, take their feedback and use it to improve your ideas iteratively. It may take time to find what subjects you like to focus on and how you like to represent ideas, and that's ok. Art, like science, is a continually evolving process, and it is important to find your own voice.

## Rule 8: Use your imagination

Readers expect comics to be imaginative and to depict ideas in new, fresh ways. A great way to communicate complex or esoteric concepts is to use analogies. Analogies allow the reader to make a connection between something that they can relate to and abstract concepts that may be complex and hard to grasp. An added benefit of analogies is that they often allow for simple variations to make a subject humorous. For example, you can equip a car with multiple "accessories" to depict the process of peer review [18] or transform a dot and circle into an acorn and a squash (Fig 8). However, be careful with analogies because they can sometimes lead to incorrect conclusions about a topic.

## Rule 9: Sketch and draft

One of the most important aspects of an effective comic is clear communication. Storyboard ideas with quick sketches. Lay out the important bits of the comic: where you want the characters, how you want the panels arranged, and where the text will go. This last point, where the text will go, is actually really important and sometimes difficult to do. Experiment with it if it doesn't seem right the first time. Choose your words. Just like a joke given by a standup comedienne, the difference between a great joke and a dud can sometimes be the specific way that you deliver it and the words that you use. You usually won't give a talk at a conference off-the-cuff, so don't do it here either! Test ideas out on others first. Write down a few ideas if you are having trouble. Sometimes the first thing that pops into your head is the best. Other times, an idea needs coaxing and refinement to really shine (Fig 9). You'll learn to recognize the difference between the two.

## Rule 10: Practice, practice, practice and have fun

No one becomes great at something instantly. Give yourself time and practice often. Sketch at conferences (see [17]), doodle during down time, and carry a notebook for ideas. Learn from others. Read some comics. There are some great ones out there and new ones popping up all the time. Find some that resonate with you and draw inspiration from them. Remember, if you have an idea, you can start without needing to do any drawing at all [19]. Use social media like Twitter, Facebook, Tumblr, and Instagram to reach your audience. Start an account for your comic and it will start to take on a life of its own! Most of all, have fun (Fig 10). Let's make that a rule.

If you are still reading, take out a piece of paper and draw a circle. Now give it some eyes and a mouth. Now have it thinking or saying something about science. Did it work? Congratulations! You are now a science comic artist!

## Acknowledgments

The authors would like to thank Philip Bourne for looking the other. . . ummm. . . helping get this kind of paper published. YB is super thankful for the NSF CAREER Award 1553289 that encourages her thinking outside the box to bring more science to more people. JEM is employed at Pacific Northwest National Laboratory (PNNL) in Richland, WA. PNNL is a multi-program national laboratory operated by Battelle for the DOE under contract DE-AC05-76RLO 1830. MP wishes to acknowledge Michelle Reeve for cartoon consultation. The authors also thank the vibrant science/art Twitter community for providing an environment to support their work.

## References

1. Concannon C. and Grenon M., Researchers: share your passion for science! Biochem Soc Trans, 2016. 44(5): p. 1507–1515. https://doi.org/10.1042/BST20160086 PMID: 27911733

2. McClain C. and Neeley L., A critical evaluation of science outreach via social media: its role and impact on scientists. F1000Res, 2014. 3: p. 300. https://doi.org/10.12688/f1000research.5918.2 PMID: 25866620
3. McClain C.R., Practices and promises of Facebook for science outreach: Becoming a "Nerd of Trust". PLoS Biol, 2017. 15(6): p. e2002020. https://doi.org/10.1371/journal.pbio.2002020 PMID: 28654674
4. Munroe, R. xkcd: A webcomic of romance, sarcasm, math, and language.; Available from: http://xkcd.com/. Accessed on 20 July 2017.
5. Munroe, R., Earth Temperature Timeline.; Available from: http://xkcd.com/1732/. Accessed on 20 July 2017.
6. Munroe, R., Old Files.; Available from: https://xkcd.com/1360/. Accessed on 20 July 2017.
7. Cham, J. PhD: Piled Higher and Deeper. Available from: http://phdcomics.com/. Accessed on 20 July 2017.
8. Partridge, M. Errant Science. Available from: http://errantscience.com/. Accessed on 20 July 2017.
9. McDermott, J. RedPen/BlackPen. Available from: http://redpenblackpen.twitter.com/. Accessed on 20 July 2017.
10. Endy D., Foundations for engineering biology. Nature, 2005. 438(7067): p. 449–53. https://doi.org/10.1038/nature04342 PMID: 16306983
11. Briscoe A.D., et al., Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genet, 2013. 9(7): p. e1003620. https://doi.org/10.1371/journal.pgen.1003620 PMID: 23950722
12. Caudron F. and Barral Y., A super-assembly of Whi3 encodes memory of deceptive encounters by single cells during yeast courtship. Cell, 2013. 155(6): p. 1244–57. https://doi.org/10.1016/j.cell.2013.10.046 PMID: 24315096
13. Krakow M., Graphic Narratives and Cancer Prevention: A Case Study of an American Cancer Society Comic Book. Health Commun, 2017. 32(5): p. 525–528. https://doi.org/10.1080/10410236.2016.1211075 PMID: 27542072
14. Tarver T., et al., A Novel tool for Health Literacy: Using Comic Books to Combat Childhood Obesity. J Hosp Librariansh, 2016. 16(2): p. 152–159. https://doi.org/10.1080/15323269.2016.1154768 PMID: 27840597
15. McNicol S., Humanising illness: presenting health information in educational comics. Med Humanit, 2014. 40(1): p. 49–55. https://doi.org/10.1136/medhum-2013-010469 PMID: 24398159
16. Thébaud O., et al., Managing marine socio-ecological systems: picturing the future. ICES J Mar Sci, 2017. fsw252.
17. Rohde M, Toselli M, and A. B. Sketchnote Army. Available from: http://sketchnotearmy.com/. Accessed on 20 September 2017.
18. McDermott J., Your Manuscript on Peer Review. Journal of Vascular and Interventional Radiology 2017. 28(5): p. 748.
19. Kessler, S. 6 Free Sites for Creating Your Own Comics. 2010; Available from: http://mashable.com/2010/10/24/create-your-own-comics/. Accessed on 20 September 2017.
20. How to Make a Comic Book. Available from: http://www.wikihow.com/Make-a-Comic-Book. Accessed on 20 September 2017.
21. D'Hont A., et al., The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature, 2012. 488(7410): p. 213–7. https://doi.org/10.1038/nature11241 PMID: 22801500
22. Harmon, D., Pilot S01E01, in Community. 2009.
23. Feldman, B. Talking to the Man Behind 'Loss,' the Internet's Longest-Running Miscarriage 'Joke'. 2015; Available from: http://nymag.com/selectall/2015/11/longest-running-miscarriage-meme-on-the-web.html. Accessed on 20 September 2017.
24. Weaver-Hightower M.B., Losing Thomas & Ella: A Father's Story (A Research Comic). J Med Humanit, 2017. 38(3): p. 215–230. https://doi.org/10.1007/s10912-015-9359-z PMID: 26463352

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for writing a response to reviewers

**William Stafford Noble***

Departments of Genome Sciences and Computer Science and Engineering, University of Washington, Seattle, Washington, United States of America

* william-noble@uw.edu

You recently submitted your first manuscript for publication, and you were pleased when the editor decided to send the manuscript out for peer review. Now you have gotten the reviews back, and the editor has asked you to revise your manuscript in light of the reviewers' comments. How should you tackle this task?

Ideally, the reviewing process can significantly improve your manuscript by allowing you to take into account the advice of multiple experts in your field. Indeed, empirical evidence suggests that papers that have undergone multiple rounds of peer review fare better in terms of citation counts than papers that are quickly accepted [1]. However, in practice, the review process can be emotionally charged as you grapple with comments that may seem to you to be ill-informed, biased, or otherwise problematic.

A well-crafted "response to reviewers" document is a critical part of your response. This document is submitted alongside your revised manuscript, summarizing the changes that you made in response to the critiques. Too frequently, authors focus on revising the manuscript itself and spend too little time making the response document clear and compelling. The result can be misunderstandings between the reviewers and the authors and ultimately, the possible rejection of a high-quality manuscript. Following are 10 simple rules that can help in formulating an effective response to reviewers.



OPEN ACCESS

**Citation:** Noble WS (2017) Ten simple rules for writing a response to reviewers. PLoS Comput Biol 13(10): e1005730. https://doi.org/10.1371/journal.pcbi.1005730

**Published:** October 12, 2017



**Funding:** The author received no specific funding for this work.

**Competing interests:** The author has declared that no competing interests exist.

## Rule 1: Provide an overview, then quote the full set of reviews

The response letter will typically begin with a summary of changes, pointing out new data and new analyses performed in response to the most essential criticisms of all the reviewers. Note that, at your discretion, the response may include figures and tables that are for the reviewers' benefit but will not go into the manuscript or supplement. These additional results can be mentioned in your Introduction. If a criticism is raised by multiple reviewers, this can also be pointed out in the summary. Thereafter, the response letter should contain the complete set of reviews with your responses interleaved.

## Rule 2: Be polite and respectful of all reviewers

Even if you are convinced that the reviewer lacks intellectual capacity, it is certainly not in your interest to convey this impression to the reviewer. Keep in mind that if the reviewer failed to understand something, the fault likely lies, at least in part, with you for not making the point clear enough. If the reviewer does not seem to be an expert in the area, remember that this level of expertise (or lack thereof) may be representative of many readers of the journal. Your goal is to make the work clear and accessible to all readers, not just to experts.

Sometimes you will need to work to understand a particular critique. In some cases, the question the reviewer asks reveals a deeper misunderstanding about the overall study or some of the assumptions therein. When specific comments seem off-base, and especially when a single reviewer has many such comments, this may be because the manuscript does not sufficiently explain the hypothesis it aims to address.

In some cases, you may believe that the reviewer is vengeful or is a competitor who has an ulterior motive to delay the manuscript. In such situations, you should not directly confront the reviewer in your response but instead communicate your concerns to the editors in a separate letter.

In rare cases, you may feel that a reviewer's critiques are simply discourteous. In such situations, it is important to remember that miscommunications are possible. Regardless, a rude critique does not justify a rude response from you, especially because your primary goal is to publish your scientific results.

## Rule 3: Accept the blame

If the reviewer failed to understand something, apologize for not making it clear. Even if you are convinced that the text is already clear (i.e., the reviewer simply missed it), consider revising the text and quoting the revised text in your response. In general, even if the requested change seems unnecessary, it is usually better to go ahead and revise with the goal of showing the reviewer that they were listened to and understood.

## Rule 4: Make the response self-contained

When you make changes to the text or to figures, quote the changes directly in the response. If possible, you can refer to the specific line number where the changes were applied, though you should be sure to specify whether you refer to the line numbers from the original or the revised manuscript. A self-contained response letter makes it easier for the reviewer to understand exactly what you did without having to flip back and forth between your manuscript and the response. Furthermore, by making your response self-contained, you reduce the likelihood that the reviewer will read the full manuscript and find new things to complain about. The only exception to this rule is when a large chunk of modified text (e.g., a new section) is too long to quote. Such changes can simply be alluded to explicitly (e.g., giving the title of the new section) in the response.

## Rule 5: Respond to every point raised by the reviewer

A frequent complaint from reviewers is that the authors failed to respond at all to several points raised in the review. In some cases, the reviewer may disagree with your response, but you should not try to avoid a difficult point by simply ignoring it.

Often, reviews will be organized into bullet points, but the reviewer may raise 2 separate issues within 1 bullet. In such situations, be sure to respond explicitly to both critiques. It is fine for you to interleave your responses in such a way that you break up 1 bullet with multiple responses. It is usually better to do this than to try to respond to multiple points in 1 block of text.

## Rule 6: Use typography to help the reviewer navigate your response

Use changes of typeface, color, and indenting to discriminate between 3 different elements: the review itself, your responses to the review, and changes that you have made to the manuscript. You can explain these typographical conventions in the introduction to your response.

## Rule 7: Whenever possible, begin your response to each comment with a direct answer to the point being raised

You can provide background information, but you should do so after giving your primary response. Provide a “yes” or “no” answer whenever possible. When the reviewer is correct, state so in your response. Your goal is to show the reviewer that you took their comments seriously, and you should quickly convey what you did in response to their critique.

## Rule 8: When possible, do what the reviewer asks

In general, you should avoid giving the impression that you couldn’t be bothered to carry out the additional experiments or analyses that the reviewer asks for. Even in cases in which you believe the reviewer has requested an analysis that you don’t find informative, or is otherwise flawed, you will often be in a stronger position if you do what the reviewer asked, report the results in your response, and then explain why you believe the results do not belong in your manuscript.

In some cases, if the reviewer makes detailed or very insightful suggestions that get incorporated into the revised manuscript, it may be appropriate to add to the Acknowledgments section an explicit "thank you" to the reviewer. Indeed, many authors routinely include an acknowledgment of the reviewers in all of their publications. Note, however, that some journals (including *PLOS Computational Biology*) do not allow reviewer acknowledgments.

Sometimes reviewers simply ask for too much. It is certainly acceptable to say that the requests go beyond what you perceive to be the scope of the current work. However, it is also important to recognize that the scope of a given manuscript is often difficult to define precisely. If the reviewer asks for 10 things, and you say that 9 out of 10 of them fall outside the scope of your work, then you are not likely to satisfy the reviewer. In such a situation, you may need to do a few things that you think fall outside the scope of your original work.

Occasionally, it may be necessary to fall back on the discretion of the editor. For example, editors often ask that authors shorten their manuscripts, whereas reviewers often ask for additional details, experiments, or analyses. If, for example, a reviewer asks you to move some content from the supplement to the main manuscript, you may want to say that you are willing to do so if the editor concurs.

## Rule 9: Be clear about what changed relative to the previous version

When you make a change in response to a reviewer’s comment, it can sometimes be difficult to convey to the reviewer exactly what that change consisted of. A common error is for an author to respond to a reviewer’s comment by saying, "This point is addressed in the manuscript in the following way. . ." This response fails to make clear whether the author is simply pointing out text that was already present in the previous version of the manuscript, or the author is describing changes that have been incorporated into the new version. In your response, refer explicitly to the previous and revised versions of your manuscript and explain what changes have been made.

## Rule 10: If necessary, write the response twice

Your initial draft of the “response to reviewers” document may aim to analyze what the reviewer meant while considering different avenues of response and the cost–benefit tradeoff of performing additional experiments. This document can be helpful to you and your coauthors as you decide how to formulate a final response document. The initial document can also be a place to vent your frustration with what you perceive to be unfair or rude reviews. After writing this initial draft, you can begin writing a completely separate document that contains what you actually want the reviewers to see. In practice, it is often helpful to write the "venting" version of the response first, wait a while, and then begin working on the "real" response several days later, perhaps after you have done some of the work to address the critiques raised by the reviewer.

In addition to the "response to reviewers" letter, you may in some cases want to write a separate letter to the managing editor. In this letter, you can address issues about potential conflicts of interest. You may also want to point out when the reviewers' requests conflict with one another or with journal policies.

The process of responding to reviewer critiques can be one of the more stressful parts of the publication process. Throughout the process, it is helpful to keep in mind that, in most cases, the reviewers are well-meaning colleagues who are volunteering their time to help ensure the validity of results that are reported in the scientific literature. In nearly every case, the manuscript that comes out of the review process is improved relative to the original version.

## Acknowledgments

The author thanks Giancarlo Bonora, Kate Cook, Max Libbrecht, Damon May, Jacob Schreiber, and Zhiping Weng for valuable feedback on an early draft of this manuscript.

## Reference

1. Calcagno V, Demoinet E, Gollner K, Guidi L, Ruths D, de Mazancourt C. Flows of research manuscripts among scientific journals reveal hidden submission patterns. Science. 2012; 338(6110):1065–1069. https://doi.org/10.1126/science.1227833 PMID: 23065906

EDITORIAL

# Ten simple rules for structuring papers

**Brett Mensh**[1,2], **Konrad Kording**[3,4]*

**1** Optimize Science, Mill Valley, California, United States of America, **2** Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, United States of America, **3** University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **4** Northwestern University, Evanston, Illinois, United States of America

* koerding@gmail.com

## Overview

Good scientific writing is essential to career development and to the progress of science. A well-structured manuscript allows readers and reviewers to get excited about the subject matter, to understand and verify the paper's contributions, and to integrate these contributions into a broader context. However, many scientists struggle with producing high-quality manuscripts and are typically untrained in paper writing. Focusing on how readers consume information, we present a set of ten simple rules to help you communicate the main idea of your paper. These rules are designed to make your paper more influential and the process of writing more efficient and pleasurable.

## Introduction

Writing and reading papers are key skills for scientists. Indeed, success at publishing is used to evaluate scientists [1] and can help predict their future success [2]. In the production and consumption of papers, multiple parties are involved, each having their own motivations and priorities. The editors want to make sure that the paper is significant, and the reviewers want to determine whether the conclusions are justified by the results. The reader wants to quickly understand the conceptual conclusions of the paper before deciding whether to dig into the details, and the writer wants to convey the important contributions to the broadest audience possible while convincing the specialist that the findings are credible. You can facilitate all of these goals by structuring the paper well at multiple scales—spanning the sentence, paragraph, section, and document.

Clear communication is also crucial for the broader scientific enterprise because "concept transfer" is a rate-limiting step in scientific cross-pollination. This is particularly true in the biological sciences and other fields that comprise a vast web of highly interconnected sub-disciplines. As scientists become increasingly specialized, it becomes more important (and difficult) to strengthen the conceptual links. Communication across disciplinary boundaries can only work when manuscripts are readable, credible, and memorable.

The claim that gives significance to your work has to be supported by data and by a logic that gives it credibility. Without carefully planning the paper's logic, writers will often be missing data or missing logical steps on the way to the conclusion. While these lapses are beyond our scope, your scientific logic must be crystal clear to powerfully make your claim.

Here we present ten simple rules for structuring papers. The first four rules are principles that apply to all the parts of a paper and further to other forms of communication such as grants and posters. The next four rules deal with the primary goals of each of the main parts of papers. The final two rules deliver guidance on the process—heuristics for efficiently constructing manuscripts.

Check for updates

OPEN ACCESS

**Citation:** Mensh B, Kording K (2017) Ten simple rules for structuring papers. PLoS Comput Biol 13(9): e1005619. https://doi.org/10.1371/journal.pcbi.1005619

**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** September 28, 2017



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Principles (Rules 1–4)

Writing is communication. Thus, the reader's experience is of primary importance, and all writing serves this goal. When you write, you should constantly have your reader in mind. These four rules help you to avoid losing your reader.

### Rule 1: Focus your paper on a central contribution, which you communicate in the title

Your communication efforts are successful if readers can still describe the main contribution of your paper to their colleagues a year after reading it. Although it is clear that a paper often needs to communicate a number of innovations on the way to its final message, it does not pay to be greedy. Focus on a single message; papers that simultaneously focus on multiple contributions tend to be less convincing about each and are therefore less memorable.

The most important element of a paper is the title—think of the ratio of the number of titles you read to the number of papers you read. The title is typically the first element a reader encounters, so its quality [3] determines whether the reader will invest time in reading the abstract.

The title not only transmits the paper's central contribution but can also serve as a constant reminder (to you) to focus the text on transmitting that idea. Science is, after all, the abstraction of simple principles from complex data. The title is the ultimate refinement of the paper's contribution. Thinking about the title early—and regularly returning to hone it—can help not only the writing of the paper but also the process of designing experiments or developing theories.

This Rule of One is the most difficult rule to optimally implement because it comes face-to-face with the key challenge of science, which is to make the claim and/or model as simple as the data and logic can support but no simpler. In the end, your struggle to find this balance may appropriately result in "one contribution" that is multifaceted. For example, a technology paper may describe both its new technology and a biological result using it; the bridge that unifies these two facets is a clear description of how the new technology can be used to do new biology.

### Rule 2: Write for flesh-and-blood human beings who do not know your work

Because you are the world's leading expert at exactly what you are doing, you are also the world's least qualified person to judge your writing from the perspective of the naïve reader. The majority of writing mistakes stem from this predicament. Think like a designer—for each element, determine the impact that you want to have on people and then strive to achieve that objective [4]. Try to think through the paper like a naïve reader who must first be made to care about the problem you are addressing (see Rule 6) and then will want to understand your answer with minimal effort.

Define technical terms clearly because readers can become frustrated when they encounter a word that they don't understand. Avoid abbreviations and acronyms so that readers do not have to go back to earlier sections to identify them.

The vast knowledge base of human psychology is useful in paper writing. For example, people have working memory constraints in that they can only remember a small number of items and are better at remembering the beginning and the end of a list than the middle [5]. Do your best to minimize the number of loose threads that the reader has to keep in mind at any one time.

## Rule 3: Stick to the context-content-conclusion (C-C-C) scheme

The vast majority of popular (i.e., memorable and re-tellable) stories have a structure with a discernible beginning, a well-defined body, and an end. The beginning sets up the context for the story, while the body (content) advances the story towards an ending in which the problems find their conclusions. This structure reduces the chance that the reader will wonder "Why was I told that?" (if the context is missing) or "So what?" (if the conclusion is missing).

There are many ways of telling a story. Mostly, they differ in how well they serve a patient reader versus an impatient one [6]. The impatient reader needs to be engaged quickly; this can be accomplished by presenting the most exciting content first (e.g., as seen in news articles). The C-C-C scheme that we advocate serves a more patient reader who is willing to spend the time to get oriented with the context. A consequent disadvantage of C-C-C is that it may not optimally engage the impatient reader. This disadvantage is mitigated by the fact that the structure of scientific articles, specifically the primacy of the title and abstract, already forces the content to be revealed quickly. Thus, a reader who proceeds to the introduction is likely engaged enough to have the patience to absorb the context. Furthermore, one hazard of excessive "content first" story structures in science is that you may generate skepticism in the reader because they may be missing an important piece of context that makes your claim more credible. For these reasons, we advocate C-C-C as a "default" scientific story structure.

The C-C-C scheme defines the structure of the paper on multiple scales. At the whole-paper scale, the introduction sets the context, the results are the content, and the discussion brings home the conclusion. Applying C-C-C at the paragraph scale, the first sentence defines the topic or context, the body hosts the novel content put forth for the reader's consideration, and the last sentence provides the conclusion to be remembered.

Deviating from the C-C-C structure often leads to papers that are hard to read, but writers often do so because of their own autobiographical context. During our everyday lives as scientists, we spend a majority of our time producing content and a minority amidst a flurry of other activities. We run experiments, develop the exposition of available literature, and combine thoughts using the magic of human cognition. It is natural to want to record these efforts on paper and structure a paper chronologically. But for our readers, most details of our activities are extraneous. They do not care about the chronological path by which you reached a result; they just care about the ultimate claim and the logic supporting it (see Rule 7). Thus, all our work must be reformatted to provide a context that makes our material meaningful and a conclusion that helps the reader to understand and remember it.

## Rule 4: Optimize your logical flow by avoiding zig-zag and using parallelism

**Avoiding zig-zag.** Only the central idea of the paper should be touched upon multiple times. Otherwise, each subject should be covered in only one place in order to minimize the number of subject changes. Related sentences or paragraphs should be strung together rather than interrupted by unrelated material. Ideas that are similar, such as two reasons why we should believe something, should come one immediately after the other.

**Using parallelism.** Similarly, across consecutive paragraphs or sentences, parallel messages should be communicated with parallel form. Parallelism makes it easier to read the text because the reader is familiar with the structure. For example, if we have three independent reasons why we prefer one interpretation of a result over another, it is helpful to communicate them with the same syntax so that this syntax becomes transparent to the reader, which allows them to focus on the content. There is nothing wrong with using the same word multiple times in a sentence or paragraph. Resist the temptation to use a different word to refer to the same concept—doing so makes readers wonder if the second word has a slightly different meaning.

## The components of a paper (Rules 5–8)

The individual parts of a paper—abstract, introduction, results, and discussion—have different objectives, and thus they each apply the C-C-C structure a little differently in order to achieve their objectives. We will discuss these specialized structures in this section and summarize them in Fig 1.

### Rule 5: Tell a complete story in the abstract

The abstract is, for most readers, the only part of the paper that will be read. This means that the abstract must convey the entire message of the paper effectively. To serve this purpose, the abstract's structure is highly conserved. Each of the C-C-C elements is detailed below.

The context must communicate to the reader what gap the paper will fill. The first sentence orients the reader by introducing the broader field in which the particular research is situated. Then, this context is narrowed until it lands on the open question that the research answered. A successful context section sets the stage for distinguishing the paper's contributions from the current state of the art by communicating what is missing in the literature (i.e., the specific gap) and why that matters (i.e., the connection between the specific gap and the broader context that the paper opened with).

The content ("Here we") first describes the novel method or approach that you used to fill the gap or question. Then you present the meat—your executive summary of the results.

Finally, the conclusion interprets the results to answer the question that was posed at the end of the context section. There is often a second part to the conclusion section that highlights how this conclusion moves the broader field forward (i.e., "broader significance"). This is particularly true for more "general" journals with a broad readership.

This structure helps you avoid the most common mistake with the abstract, which is to talk about results before the reader is ready to understand them. Good abstracts usually take many iterations of refinement to make sure the results fill the gap like a key fits its lock. The broad-narrow-broad structure allows you to communicate with a wider readership (through breadth) while maintaining the credibility of your claim (which is always based on a finite or narrow set of results).

### Rule 6: Communicate why the paper matters in the introduction

The introduction highlights the gap that exists in current knowledge or methods and why it is important. This is usually done by a set of progressively more specific paragraphs that culminate in a clear exposition of what is lacking in the literature, followed by a paragraph summarizing what the paper does to fill that gap.

As an example of the progression of gaps, a first paragraph may explain why understanding cell differentiation is an important topic and that the field has not yet solved what triggers it (a field gap). A second paragraph may explain what is unknown about the differentiation of a specific cell type, such as astrocytes (a subfield gap). A third may provide clues that a particular gene might drive astrocytic differentiation and then state that this hypothesis is untested (the gap within the subfield that you will fill). The gap statement sets the reader's expectation for what the paper will deliver.

The structure of each introduction paragraph (except the last) serves the goal of developing the gap. Each paragraph first orients the reader to the topic (a context sentence or two) and then explains the "knowns" in the relevant literature (content) before landing on the

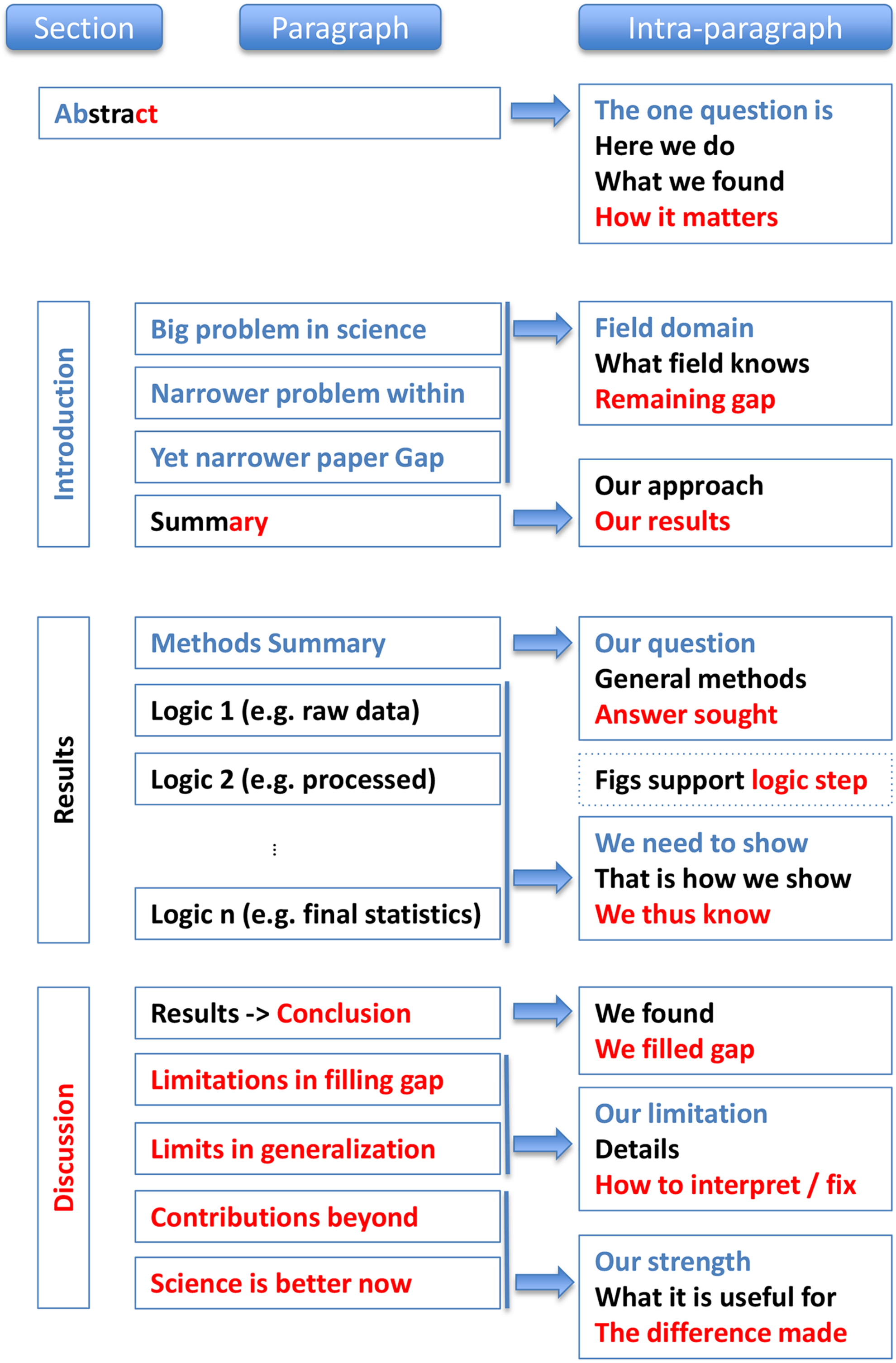



**Fig 1. Summary of a paper's structural elements at three spatial scales: Across sections, across paragraphs, and within paragraphs.** Note that the abstract is special in that it contains all three elements (Context, Content, and Conclusion), thus comprising all three colors.

https://doi.org/10.1371/journal.pcbi.1005619.g001

critical "unknown" (conclusion) that makes the paper matter at the relevant scale. Along the path, there are often clues given about the mystery behind the gaps; these clues lead to the untested hypothesis or undeveloped method of the paper and give the reader hope that the mystery is solvable. The introduction should not contain a broad literature review beyond the motivation of the paper. This gap-focused structure makes it easy for experienced readers to evaluate the potential importance of a paper—they only need to assess the importance of the claimed gap.

The last paragraph of the introduction is special: it compactly summarizes the results, which fill the gap you just established. It differs from the abstract in the following ways: it does not need to present the context (which has just been given), it is somewhat more specific about the results, and it only briefly previews the conclusion of the paper, if at all.

## Rule 7: Deliver the results as a sequence of statements, supported by figures, that connect logically to support the central contribution

The results section needs to convince the reader that the central claim is supported by data and logic. Every scientific argument has its own particular logical structure, which dictates the sequence in which its elements should be presented.

For example, a paper may set up a hypothesis, verify that a method for measurement is valid in the system under study, and then use the measurement to disprove the hypothesis. Alternatively, a paper may set up multiple alternative (and mutually exclusive) hypotheses and then disprove all but one to provide evidence for the remaining interpretation. The fabric of the argument will contain controls and methods where they are needed for the overall logic.

In the outlining phase of paper preparation (see Rule 9), sketch out the logical structure of how your results support your claim and convert this into a sequence of declarative statements that become the headers of subsections within the results section (and/or the titles of figures). Most journals allow this type of formatting, but if your chosen journal does not, these headers are still useful during the writing phase and can either be adapted to serve as introductory sentences to your paragraphs or deleted before submission. Such a clear progression of logical steps makes the paper easy to follow.

Figures, their titles, and legends are particularly important because they show the most objective support (data) of the steps that culminate in the paper's claim. Moreover, figures are often viewed by readers who skip directly from the abstract in order to save time. Thus, the title of the figure should communicate the conclusion of the analysis, and the legend should explain how it was done. Figure making is an art unto itself; the Edward Tufte books remain the gold standard for learning this craft [7,8].

The first results paragraph is special in that it typically summarizes the overall approach to the problem outlined in the introduction, along with any key innovative methods that were developed. Most readers do not read the methods, so this paragraph gives them the gist of the methods that were used.

Each subsequent paragraph in the results section starts with a sentence or two that set up the question that the paragraph answers, such as the following: "To verify that there are no artifacts. . .," "What is the test-retest reliability of our measure?," or "We next tested whether $Ca^{2+}$ flux through L-type $Ca^{2+}$ channels was involved." The middle of the paragraph presents data and logic that pertain to the question, and the paragraph ends with a sentence that answers the question. For example, it may conclude that none of the potential artifacts were detected. This structure makes it easy for experienced readers to fact-check a paper. Each paragraph convinces the reader of the answer given in its last sentence. This makes it easy to find the paragraph in which a suspicious conclusion is drawn and to check the logic of that paragraph. The result of each paragraph is a logical statement, and paragraphs farther down in the text rely on the logical conclusions of previous paragraphs, much as theorems are built in mathematical literature.

### Rule 8: Discuss how the gap was filled, the limitations of the interpretation, and the relevance to the field

The discussion section explains how the results have filled the gap that was identified in the introduction, provides caveats to the interpretation, and describes how the paper advances the field by providing new opportunities. This is typically done by recapitulating the results, discussing the limitations, and then revealing how the central contribution may catalyze future progress. The first discussion paragraph is special in that it generally summarizes the important findings from the results section. Some readers skip over substantial parts of the results, so this paragraph at least gives them the gist of that section.

Each of the following paragraphs in the discussion section starts by describing an area of weakness or strength of the paper. It then evaluates the strength or weakness by linking it to the relevant literature. Discussion paragraphs often conclude by describing a clever, informal way of perceiving the contribution or by discussing future directions that can extend the contribution.

For example, the first paragraph may summarize the results, focusing on their meaning. The second through fourth paragraphs may deal with potential weaknesses and with how the literature alleviates concerns or how future experiments can deal with these weaknesses. The fifth paragraph may then culminate in a description of how the paper moves the field forward. Step by step, the reader thus learns to put the paper's conclusions into the right context.

## Process (Rules 9 and 10)

To produce a good paper, authors can use helpful processes and habits. Some aspects of a paper affect its impact more than others, which suggests that your investment of time should be weighted towards the issues that matter most. Moreover, iteratively using feedback from colleagues allows authors to improve the story at all levels to produce a powerful manuscript. Choosing the right process makes writing papers easier and more effective.

### Rule 9: Allocate time where it matters: Title, abstract, figures, and outlining

The central logic that underlies a scientific claim is paramount. It is also the bridge that connects the experimental phase of a research effort with the paper-writing phase. Thus, it is useful to formalize the logic of ongoing experimental efforts (e.g., during lab meetings) into an evolving document of some sort that will ultimately steer the outline of the paper.

You should also allocate your time according to the importance of each section. The title, abstract, and figures are viewed by far more people than the rest of the paper, and the methods section is read least of all. Budget accordingly.

The time that we do spend on each section can be used efficiently by planning text before producing it. Make an outline. We like to write one informal sentence for each planned paragraph. It is often useful to start the process around descriptions of each result—these may become the section headers in the results section. Because the story has an overall arc, each paragraph should have a defined role in advancing this story. This role is best scrutinized at the outline stage in order to reduce wasting time on wordsmithing paragraphs that don't end up fitting within the overall story.

**Table 1. A summary of the ten rules and how to tell if they are being violated.**

| Rule | Sign it is violated |
|---|---|
| 1: Focus on one big idea | Readers cannot give 1-sentence summary. |
| 2: Write for naive humans | Readers do not "get" the paper. |
| 3: Use context, content, conclusion structure | Readers ask why something matters or what it means. |
| 4: Optimize logical flow | Readers stumble on a small section of the text. |
| 5: Abstract: Compact summary of paper | Readers cannot give the "elevator pitch" of your work after reading it. |
| 6: Introduction: Why the paper matters | Readers show little interest in the paper. |
| 7: Results: Why the conclusion is justified | Readers do not agree with your conclusion. |
| 8: Discussion: Preempt criticism, give future impact | Readers are left with unanswered criticisms and/or questions on their mind. |
| 9: Allocate time wisely | Readers struggle to understand your central contribution despite your having worked hard. |
| 10: Iterate the story | The paper's contribution is rejected by test readers, editors, or reviewers. |

https://doi.org/10.1371/journal.pcbi.1005619.t001

## Rule 10: Get feedback to reduce, reuse, and recycle the story

Writing can be considered an optimization problem in which you simultaneously improve the story, the outline, and all the component sentences. In this context, it is important not to get too attached to one's writing. In many cases, trashing entire paragraphs and rewriting is a faster way to produce good text than incremental editing.

There are multiple signs that further work is necessary on a manuscript (see Table 1). For example, if you, as the writer, cannot describe the entire outline of a paper to a colleague in a few minutes, then clearly a reader will not be able to. You need to further distill your story. Finding such violations of good writing helps to improve the paper at all levels.

Successfully writing a paper typically requires input from multiple people. Test readers are necessary to make sure that the overall story works. They can also give valuable input on where the story appears to move too quickly or too slowly. They can clarify when it is best to go back to the drawing board and retell the entire story. Reviewers are also extremely useful. Non-specific feedback and unenthusiastic reviews often imply that the reviewers did not "get" the big picture story line. Very specific feedback usually points out places where the logic within a paragraph was not sufficient. It is vital to accept this feedback in a positive way. Because input from others is essential, a network of helpful colleagues is fundamental to making a story memorable. To keep this network working, make sure to pay back your colleagues by reading their manuscripts.

## Discussion

This paper focused on the structure, or "anatomy," of manuscripts. We had to gloss over many finer points of writing, including word choice and grammar, the creative process, and collaboration. A paper about writing can never be complete; as such, there is a large body of literature dealing with issues of scientific writing [9,10,11,12,13,14,15,16,17].

Personal style often leads writers to deviate from a rigid, conserved structure, and it can be a delight to read a paper that creatively bends the rules. However, as with many other things in life, a thorough mastery of the standard rules is necessary to successfully bend them [18]. In following these guidelines, scientists will be able to address a broad audience, bridge disciplines, and more effectively enable integrative science.

## Acknowledgments

We took our own advice and sought feedback from a large number of colleagues throughout the process of preparing this paper. We would like to especially thank the following people who gave particularly detailed and useful feedback:

Sandra Aamodt, Misha Ahrens, Vanessa Bender, Erik Bloss, Davi Bock, Shelly Buffington, Xing Chen, Frances Cho, Gabrielle Edgerton, multiple generations of the COSMO summer school, Jason Perry, Jermyn See, Nelson Spruston, David Stern, Alice Ting, Joshua Vogelstein, Ronald Weber.

## References

1. Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 102: 16569–16572. https://doi.org/10.1073/pnas.0507655102 PMID: 16275915
2. Acuna DE, Allesina S, Kording KP (2012) Future impact: Predicting scientific success. Nature. 489: 201–202. https://doi.org/10.1038/489201a PMID: 22972278
3. Paiva CE, Lima JPSN, Paiva BSR (2012) Articles with short titles describing the results are cited more often. Clinics. 67: 509–513. https://doi.org/10.6061/clinics/2012(05)17 PMID: 22666797
4. Carter M (2012) Designing Science Presentations: A Visual Guide to Figures, Papers, Slides, Posters, and More: Academic Press.
5. Murdock BB Jr (1968) Serial order effects in short-term memory. J Exp Psychol. 76: Suppl:1–15.
6. Schimel J (2012) Writing science: how to write papers that get cited and proposals that get funded. USA: OUP.
7. Tufte ER (1990) Envisioning information. Graphics Press.
8. Tufte ER The Visual Display of Quantitative Information. Graphics Press.
9. Lisberger SG (2011) From Science to Citation: How to Publish a Successful Scientific Paper. Stephen Lisberger.
10. Simons D (2012) Dan's writing and revising guide. http://www.dansimons.com/resources/Simons_on_writing.pdf [cited 2017 Sep 9].
11. Sørensen C (1994) This is Not an Article—Just Some Thoughts on How to Write One. Syöte, Finland: Oulu University, 46–59.
12. Day R (1988) How to write and publish a scientific paper. Phoenix: Oryx.
13. Lester JD, Lester J (1967) Writing research papers. Scott, Foresman.
14. Dumont J-L (2009) Trees, Maps, and Theorems. Principiae. http://www.treesmapsandtheorems.com/ [cited 2017 Sep 9].
15. Pinker S (2014) The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century. Viking Adult.
16. Bern D (1987) Writing the empirical journal. The compleat academic: A practical guide for the beginning social scientist. 171.
17. George GD, Swan JA (1990) The science of scientific writing. Am Sci. 78: 550–558.
18. Strunk W (2007) The elements of style. Penguin.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules to consider regarding preprint submission

**Philip E. Bourne[1] *, Jessica K. Polka[2], Ronald D. Vale[3], Robert Kiley[4]**

**1** Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America, **2** Whitehead Institute, Cambridge, Massachusetts, United States of America, **3** Department of Cellular and Molecular Pharmacology and the Howard Hughes Medical Institute, University of California San Francisco, San Francisco, California, United States of America, **4** Wellcome Library, The Wellcome Trust, London, United Kingdom

* pebourne@gmail.com

Check for updates

OPEN ACCESS

**Citation:** Bourne PE, Polka JK, Vale RD, Kiley R (2017) Ten simple rules to consider regarding preprint submission. PLoS Comput Biol 13(5): e1005473. https://doi.org/10.1371/journal.pcbi.1005473

**Published:** May 4, 2017



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

For the purposes of these rules, a preprint is defined as a complete written description of a body of scientific work that has yet to be published in a journal. Typically, a preprint is a research article, editorial, review, etc. that is ready to be submitted to a journal for peer review or is under review. It could also be a commentary, a report of negative results, a large data set and its description, and more. Finally, it could also be a paper that has been peer reviewed and either is awaiting formal publication by a journal or was rejected, but the authors are willing to make the content public. In short, a preprint is a research output that has not completed a typical publication pipeline but is of value to the community and deserving of being easily discovered and accessed. We also note that the term preprint is an anomaly, since there may not be a print version at all. The rules that follow relate to all these preprint types unless otherwise noted.

In 1991, physics (and later, other disciplines, including mathematics, computer science, and quantitative biology) began a tradition of making preprints available through arXiv [1]. arXiv currently contains well over 1 million preprints. While late to the game [2], the availability of preprints in biomedicine has gained significant community attention recently [3,4] and led to the formation of a scientist-driven effort, ASAPbio [5], to promote their use. As a result of an ASAPbio meeting held in February of 2016, a paper was published [6] that describes the pros and cons of preprints from the perspective of the stakeholders—scientists, publishers, and funders. Here, we formulate the message specifically for scientists in the form of ten simple rules for considering using preprints as a communication mechanism.

## Rule 1: Preprints speed up dissemination

A recent analysis highlighted that the median review time—the time between submission and acceptance of an article—is around 100 days, with a further 25 days or so spent preparing the work for publication [7]. However, these figures—slow as they are—do not include the time researchers spend "shopping around" for a journal to publish their findings, which can induce rounds of editorial rejection before or after peer review. Stephen Royle, a cell biologist at the University of Warwick, undertook an analysis of his published papers over the past dozen years and concluded that the average time from first submission to publication was around 9 months [8]. Royle's is one example of a well-studied phenomenon [9]. In summary, at a time when technology allows research findings to be shared instantly, the time to access research output appears glacial and similar to the pre-internet era.

## Rule 2: Preprints should be licensed and formatted to facilitate reuse

In principle, preprints can be text and data mined to better comprehend and utilize the knowledge presented. This assumes that copyright, licensing, and format permit such use. Maximizing accessibility and reuse is not necessarily the default currently offered by preprint services. Consequently, when posting a preprint, authors are encouraged to use licenses and formats that facilitate reuse while retaining copyright to their work. Details of copyright, licensing, and format are beyond the scope of this article, but licensing your work as CC-BY (reusable by all, provided attribution is given) and providing a text-accessible version covers most situations. Software tools that facilitate the comprehension of accessible content (for example, Content Mine) are in their infancy but are likely to become mainstream in the next 5–10 years. Better still is the promise that the traditional content of research articles can be integrated with the underlying data, analytics, and commentary to create a new learning experience. To the community, this represents an opportunity to accelerate discovery in ways that are not currently offered by traditional publishers to the contributing authors. Such an offering would presumably provide new opportunities for an author's work to be used and cited.

## Rule 3: Preprints provide a record of priority

There are a number of resources that provide preprint services to the biosciences (for example, bioRxiv [10], PeerJ Preprints [11], and the Quantitative Biology section within arXiv [12]). All include an uneditable timestamp indicating when the article appeared, which is usually within 24 hours of submission. This date, along with the preprint itself, is made open access (see Rule 2), and thus, anyone (using any internet search engine) can determine the order of priority relative to other published work or, indeed, other preprints. One of the original motivations for creating arXiv was to create a transparent public record of a scientist's work. By contrast, while journals provide an important service of validation through peer review, establishment of priority can be significantly delayed because the work is not public during the process of peer review in most journals.

The complementary roles of preprints and journals in establishing priority and validation, respectively, are discussed in a commentary by Vale and Hyman [13]. Since preprints may extend beyond traditional published papers, they create an order of priority for these research products as well.

## Rule 4: Preprints do not lead to being scooped

Many scientists wonder if they might be scooped if their work is made public ahead of the formal journal publication. Stepping back, perhaps we should ask: what is the definition of scooping? Here, we take it to mean that, either inadvertently or purposely, an author publishes a biomedical finding and does not provide attribution to the original author(s). The notion that preprints leads to scooping is covered in some detail by ASAPbio [14], and only a synopsis is given here. Again, the presence of arXiv provides a history of what has happened, at least in other disciplines. The short answer, according to Paul Ginsparg, the creator of arXiv, is that intentional scooping is virtually absent in physics because these scientists are aware of the arXiv communication and do not tolerate such behavior. Then, the question becomes whether the biomedical community is somehow different in its ethics or behavior. We believe not, and there is no evidence that this is happening with current preprints. Furthermore, as preprints become more visible and commonplace (like arXiv), scooping will be become increasingly difficult. By contrast, with a nonpublic publication process, it is hard for authors to prove originality during this period if nothing about the work is registered in the public domain. Posters and oral presentations might prove originality, but they are often not publicly and persistently available or detailed enough to support the originality of a body of work. Preprints address this issue, as described in Rule 3, and they can and should be fairly cited.

## Rule 5: Preprints provide access to scholarly content that would otherwise be lost

In addition to our formal publications, as scientists, we have scholarly outputs that we are willing to stand behind but may not have an outlet: a graduate student leaves, gets tied up in a new position, and the paper never gets that final polish yet contains meaningful results and conclusions; a project yields negative data or data that simply does not come together into a coherent story yet has value to the community; replication of a study (or not) represents a useful outcome but is not innovative enough for journal publication. In summary, preprints offer a way of sharing important scholarly output that would otherwise disappear after much time and expense.

Some might argue that work that has not passed peer review should be disregarded. To those, we say, "How much useful information do you get from discussions of unpublished data at meetings, in blogs, and via other forms of non-peer-reviewed content?" We would argue that this type of useful information is growing in both volume and importance. The same naysayers will then likely say, "There is too much misinformation as well as useful information on the internet." We agree that filters are needed. Human filters will not be able to cope with the volume, hence the need for software tools as described in Rule 2.

## Rule 6: Preprints do not imply low quality

Given that preprints have not been peer reviewed, does that imply low quality? Certainly, the peer review process can add significant value to the work, pointing out errors or areas for improvement. Nevertheless, authors must stand behind their submitted preprint, because it is a public disclosure (and hence a citable entity), albeit a non-peer-reviewed one. Even without peer review, their scientific colleagues will be reading and judging the work, and the authors' reputations are at stake. Thus, scientists will be careful to disclose their best work that reflects their scientific abilities and expertise, so work of low quality would not be expected. This has been true of arXiv over the years, and the high-quality factor also seems to apply to bioRxiv [10]. To illustrate this, we know a high-profile biomedical research laboratory that now conducts their journal clubs exclusively on preprints [15].

## Rule 7: Preprints support the rapid evaluation of controversial results

Science is, by its nature, iterative and self-correcting. Through preprints, the time to correction can be much reduced. Experience with arXiv has shown that claims concerning, for example, superluminal neutrinos [16] or bicep2 primordial gravitational waves [17] could be discredited before they reached the published literature. In biomedicine, a case in point was the publication of information in May of 2016 [18] that indicated cell phone radiation boosts cancer rates in animals. Given the controversy around such a statement, the National Institutes of Health (NIH) felt an obligation to release all the data, including internal reviews, as quickly as possible so that others could review the findings. This would not be possible through conventional publishing, since neither the form of the manuscript nor the inclusion of an internal review would be suitable for most journals, but a preprint [19] was posted within 24 hours. In a little over 5 months since the preprint was posted, it has been downloaded 148,000 times, providing a more complete picture of the controversial result. It could be argued that the preprint furthered the controversy, but it could also be argued that the authors were under an obligation to provide all available data to describe the research. You could take this further and argue that the science should have been open as it progressed, but that is still not within the comfort zone of most scientists.

## Rule 8: Preprints do not typically preclude publication

Sherpa/Romeo [20] tracks the preprint policies of publishers and their associated academic journals. As can be seen there (and further outlined by [9]), very few journals consider preprints as a “prior form of publication” and reject such manuscripts on the grounds that they had been posted to a preprint server. This is in contrast to the Ingelfinger Rule, enunciated in 1969 by Franz J. Ingelfinger on behalf of the New England Journal of Medicine [21] and followed by many other journals, that would not publish material made available in other media or in other journals. Today, journals publishing papers that have appeared as preprints either speaks to a relaxation of the so-called Ingelfinger Rule or to the idea that preprints are not considered prior publication. In any case, in recent months, more life science journals are developing preprint-friendly policies—and a number have mechanisms to accept journal submissions directly from bioRxiv [16]. We expect this trend to continue as publishers grow to appreciate the value of preprints and how community input can help the author to improve their work and manuscript, leading to a better publication of record.

## Rule 9: Preprints can further inform grant review and academic advancement

The lack of a substantive body of work in support of a particular grant application or academic promotion can be a substantial obstacle to career advancement, particularly for young scientists.

First, consider grant applications to funding bodies. Papers submitted (or even accepted) but not yet published do not help, since the grant reviewer cannot judge the work. By contrast, the availability of preprints can provide a reviewer with the evidence they need to substantiate recent productivity, as well as support the work being proposed in the grant application. It can be argued that this creates more work for the reviewer, but this work results in the ability to perform a more informed review. How individual funders currently treat preprints is variable, and thus, their value to scientists in the way described is also variable. NIH has recently encouraged the inclusion of preprints in grant applications and reports [22]. The Wellcome Trust supports the inclusion of preprints in grant applications and end-of-grant reports [23], the Simons Foundation encourages scientists to post preprints [24], and the Human Frontiers Science Program will allow them to be listed on applications and reports starting in 2017 [25]. Likewise, the Medical Research Council (MRC UK) [26], the Helmsley Charitable Trust [27], and the Canadian Institutes for Health Research [28] are actively encouraging preprints. Currently, many funding agencies are reevaluating their policies (or lack of policies) regarding preprints, so we expect many new pro-preprint policies to emerge in the coming year. Progress of funders in this regard can be tracked from the ASAPbio website [29].

Now consider academic advancement. At the time of academic promotion, a significant body of a scientist’s work could be tied up in the journal review and publication pipeline. Certainly, submitted papers can usually form part of a promotion file, but this carries less weight and credibility than a preprint, which is an acknowledgment by the author that the work is worthy of public viewing and dissemination to the entire scientific community. Moreover, if a knowledgeable reader has significant thoughts on the preprint, those could be posted as commentary, at least on some preprint services. This has wider ramifications, since commentary on preprints may provide the opportunity to improve the final published paper.

## Rule 10: Preprints—one shoe does not fit all

bioRxiv, which is the fastest-growing preprint repository for the life sciences, does not accept preprints that, if posted, could have a damaging effect on human health. This makes sense. Since submissions to bioRxiv only undergo a cursory human review before being posted, there is the possibility that potentially harmful information (e.g., unverified claims about the side effects of vaccines, etc.) or perhaps private and personal information may be revealed. This has ethical, legal, and social issues (ELSI). Such arguments flow into issues of intellectual property (IP) associated with the content of a preprint (noting that IP runs counter to Rule 2), wherein there is the risk of undesirable public release of information. It should be noted that this is not an issue restricted to preprints but one that can apply to talks, posters, etc. too. For research articles, professional editors and reviewers provide additional layers to safeguard from sensitive content being inadvertently released. Currently, preprints have only cursory safeguards, though a future preprint service could enable more rigorous review.

With open content from preprint services available through application program interfaces (APIs), there is the exciting opportunity for researchers to develop tools to better automatically or semi-automatically flag potential ELSI and IP issues. If those tools were open, they would benefit the publishing industry as well.

What should be apparent from these ten simple rules is that the provision and use of preprints in the biomedical sciences is still evolving, but there are clear benefits to the individual and the community. ASAPbio is in the process of developing a governance structure that includes all stakeholders to recommend how best to move forward with the further use of preprints. We invite you to contribute your next paper as a preprint and join the movement.

The original version of this article, prior to peer review, can be found as a preprint here [30].

## References


1. arXiv.org e-Print archive. https://arxiv.org/. Accessed 31 March 2017.
2. Ginsparg P. Preprint Deja vu The EMBO Journal 2016
3. Lauer MS, Krumholz HM, Topol EJ. Time for a prepublication culture in clinical research? Lancet 2015.
4. Vale RD Accelerating scientific publication in biology PNAS 2015
5. ASAPbio Accelerating Science and Publication in Biology. http://asapbio.org/. Accessed 31 March 2017.
6. Berg JM, Bhalla N, Bourne PE, Chalfie M, Drubin DG, Fraser JS, Greider CW, Hendricks M, Jones C, Kiley R, King S, Kirschner MW, Krumholtz HM, Lehmann R, Leptin M, Pulverer B, Rosenzweig B, Spiro JE, Stebbins M, Strasser C, Swaminathan S, Turner P, Vale RD, VijayRaghavan K, Wolberger Science 2016
7. Powell K Does it take too long to publish research? Nature 2016
8. Some Things Last a Long Time. (2014 April 15). https://quantixed.wordpress.com/2014/04/15/some-things-last-a-long-time/. Accessed 31 March 2017.
9. Borgman CL 2007 Scholarship in the Digital Age: Information, Infrastructure and the Internet MIT Press, Cambridge MA.
10. bioRxiv The Preprint Server for Biology. http://biorxiv.org/. Accessed 31 March 2017.
11. Preprints (not yet peer-reviewed). https://peerj.com/preprints-search/. Accessed 31 March 2017.
12. Quantitative Biology (since Sep. 2003). https://arxiv.org/archive/q-bio. Accessed 31 March 2017.

13. Vale RD, Hyman AA 2016 Point or view: Priority of discovery in the life sciences eLife http://dx.doi.org/10.7554/eLife.16931.
14. ArXiv founder Paul Ginsparg's thoughts on scooping. http://asapbio.org/preprint-info/preprint-faq#qe-faq-923. Accessed 31 March 2017.
15. Open-content preprint peer review from John Didion. http://academickarma.org/openreviews?id=@jdidion. Accessed 31 March 2017.
16. Faster-than-light neutrino anomaly. (2017 March 20). https://en.wikipedia.org/wiki/Faster-than-light_neutrino_anomaly. Accessed 31 March 2017.
17. Cowen R. Gravitational waves study now officially dead Nature 2015
18. Cell Phone Radiation Boosts Cancer Rates in Animals; $25Million NTP Study Finds Brain Tumors. (2016 May 25). http://microwavenews.com/news-center/ntp-cancer-results. Accessed 31 March 2017.
19. Wyde M, Cesta M, Blystone C, Elmore S, Foster P, Hooth M, et al. Report of partial findings from the national toxicology program carcinogenesis studies of cell phone radiofrequency radiation in Hsd Sprague Dawley rats (whole body exposure) (2016 June 23) http://biorxiv.org/content/early/2016/06/23/055699. Accessed 31 March 2017.
20. ShERPA/RoMEO. http://www.sherpa.ac.uk/romeo/index.php. Accessed 31 March 2017.
21. Ingelfinger F Definition of "sole contribution" New England Journal of Medicine 281, 676–677. https://doi.org/10.1056/NEJM196909182811208 PMID: 5807917
22. Reporting preprints and other interim research products (2017 March 24) https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html. Accessed 17 April 2017.
23. Simons Foundation (2016, September 1). http://simonsfoundation.s3.amazonaws.com/share/Policies_and_forms/2017PilotAndResearchAwards%20/Simons_Foundation_Policy_and_Procedures.pdf. Accessed 31 March 2017.
24. We now accept preprints in grant applications. (2017 January 10). https://wellcome.ac.uk/news/we-now-accept-preprints-grant-applications. Accessed 31 March 2017.
25. Human Frontiers Science Program Use of Preprint Servers. http://www.hfsp.org/funding/use-preprint-servers. Accessed 31 March 2017.
26. The MRC supports preprints. (2017 January 3). https://www.mrc.ac.uk/news/browse/the-mrc-supports-preprints/. Accessed 31 March 2017.
27. Helmsley encourages preprints for biomedical research grantees (2016 August 29). http://helmsleytrust.org/news/helmsley-encourages-preprints-biomedical-research-grantees. Accessed 31 March 2017.
28. CIHR peer review manual for grant applications. (2016 August 19). http://www.cihr-irsc.gc.ca/e/4656.html. Accessed 31 March 2017.
29. ASAPbio funder policies. http://asapbio.org/funder-policies. Accessed 31 March 2017.
30. Bourne PE, Polka JK, Vale RD, Kiley R. 2016 PeerJ Preprints 4:e2669v1 https://doi.org/10.7287/peerj.preprints.2669v1

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for short and swift presentations

**Christopher J. Lortie**[1,2]*

**1** Department of Biology, York University, Toronto, Ontario, Canada, **2** National Center for Ecological Analysis and Synthesis, UCSB, Santa Barbara, California, United States of America

* lortie@yorku.ca

OPEN ACCESS

**Citation:** Lortie CJ (2017) Ten simple rules for short and swift presentations. PLoS Comput Biol 13(3): e1005373. https://doi.org/10.1371/journal.pcbi.1005373

**Published:** March 30, 2017



**Funding:** The author received no specific funding for this work.

**Competing interests:** The author has declared that no competing interests exist.

## Preamble

Scientific communication is an independent research domain and has become a fundamental component of most scientific discourse and all public outreach. It now comprises a set of critical activities for many research programs [1, 2], including those that directly influence global and human health [3]. Scientific communication has evolved because it does not have to happen only at the final stages of a research endeavor but can be used to engage the public to fund the research (https://scifundchallenge.org), participate in the data collection (http://www.audubon.org/conservation/science/christmas-bird-count), share or crowd source the code and analyses (https://github.com), and process the evidence (https://www.zooniverse.org). Unfortunately, scientific progress in some fields such as climate change has outpaced our capacity to effectively communicate and contextualize findings for the public [4]. To mitigate this shortcoming, resources specific to scientists have been developed [5–8]. Boot-camp training workshops are now also offered (i.e., American Institute of Biological Sciences [AIBS]), and discussion of how academics use brief communications, such as social media tools, is present within the primary research literature [9–11]. An interesting related opportunity has emerged that, in some respects, bridges the gap between lengthy, detailed presentations of scientific findings and "sound bites" such as headlines or short press releases appropriate for media reporting: very short, swift presentations. Admittedly, these talks are in essence sound bites, too, but with more depth and thus requiring special consideration in terms of how to best leverage their potential [12]. These shorter presentations are commonly directed both to peers at scientific conferences and to the general public at in-person events and online. This format is particularly suited to online dissemination and sharing through YouTube, with most major scientific institutions and organizations administering channels of curated content. Many major scientific conventions include offerings of rapid-fire format talks—at first to communicate meta-science but now also to share primary research findings. The specific guidelines vary, but the slide deck is often limited by a set number of slides, or the presentation is limited by very strict, short time constraints (such as found with lightning talks). In addition, the slides can be set to rapidly autoadvance, for instance, with PechaKucha presentations. These presentation formats are also organized into open, public series and feature involvement from experts in many disciplines on numerous topics, including science. Succinct prose is thus a critical element in communicating science using these presentation formats. On a cautionary note, reducing much longer talks to these shorter formats is likely not the most effective strategy because shorter total presentation times coupled with rapid pacing can dramatically influence the scope and depth of the material. Best practices for scientific communication certainly apply to these talks, but specific strategies are nonetheless needed. For instance, as a general rule-of-thumb, talks prepared for a more general public audience should emphasize the implications of the science and use direct, natural language and visual analogies (instead of necessarily always showing complex evidence or primary data). Talks for scientific colleagues must also embrace parsimony but can accommodate more technical language depending on the specific audience, more direct evidence, and some data visualization that highlights complexity appropriately.

Effective oral prose is not dissimilar from effective writing. Depending on the literary theory and school of criticism that one subscribes, concepts such as "lightness, quickness, exactitude, visibility, multiplicity," and also consistency in writing [13] similarly apply to rapid presentations. The simple rules for making good presentations also apply [14], but short, swift presentations provide both novel challenges and unique communication opportunities. The pace is rapid, providing very limited time for the audience to read or process an individual slide. Thus, Rule 8 from a previous ten simple rules for oral presentations, "use visuals sparingly but effectively" [14], is a best practice for this context. If the slide deck autoadvances, the speaker must perfectly time delivery, and thus, preparation relative to a longer, less structured talk differs. The net time is significantly reduced from even most typical conference oral presentations, thereby limiting the potential scope of coverage of a topic and depth. This suggests a further reduction in the number of take-home messages suggested for longer presentations (i.e., Rule 4 from a previous set of rules suggests three points should be retained, whilst here it is likely fewer, such as one) [14]. These challenges can become benefits if handled effectively. A swift tempo engenders enthusiasm, energy, and the expectation that a bird's-eye view of a topic will be provided to, quite literally, "get the audience up to speed" on the salient issues. In the spirit of light, quick, and exact (but without the lazy dog), here are ten simple rules for presentation formats that do not wait for the speaker. A slide deck and a video of these suggestions are also available (http://bit.ly/short-swift).

## Rule 1: Plan a clear story

Avoid detours, tangents, or side anecdotes. Amusing anecdotes by accomplished speakers can be compelling, useful tools to engage and connect the speaker to the audience emotionally. In longer talks, they can also serve as a reprieve from detail-laden or inaccessible issues, and anecdotes can reframe the science into more general contexts. In a short, brisk talk, however, immediacy is paramount, and tangents are best avoided. Prepare one primary message for the audience. A total of 20 slides or 5–6 minutes, for instance, do not leave sufficient room for a story within a story. A clear story can captivate and illuminate, but a planned story is more likely to do both.

## Rule 2: Provide only one major point per slide

You have a story to tell with very limited time. Ensure each slide is a meaningful step. Some of the slides can be used to support a difficult step taken by clarifying the briefest of introductions on the previous slide. This reinforcement ensures that the audience is sufficiently informed to move forward with you on the following slide. Build on your clear story (Rule 1) one step at a time. Balance support and advancement appropriately.

## Rule 3: Limit use of text

It is much quicker for you to directly state the purpose of a given slide. Nonetheless, parsimonious use of text can assist the audience in scanning each slide for meaning and relevance. Treat the slides like scientific figures—"captions are not optimal" [15] but can be powerful aids if they do not detract from the visuals. An alternative approach is to show a visual/figure on a full slide, maximized for viewing, and use the subsequent slide for a single, brief sentence stating the finding or implication. This has added value in that it provides the speaker with more time to explain the findings and mimics a rapid but effective show-and-tell approach. Important data visualization can benefit from this presentation technique. This is a specific strategy that can work for some but not all. The overarching principle is that an effective talk balances text with visuals and oral explanation. One must provide enough to read but not overwhelm so as to avoid the audience hurriedly reading throughout the presentation. Better they pay attention to you than to your slide deck.

## Rule 4: Use simple visuals

Slides advance very rapidly in these talk formats. Similar to the rules for better figures [15], identify the key message and avoid superfluous visual elements. Do not cut and paste figures prepared for written papers because the risk of losing the audience in a rapid talk is too great if they are expected to search, parse, or mentally rotate elements such as labels. Simplify data visualizations as needed and use color to show groupings and patterns. Visual guides and color are allowed here and not necessarily bound by the same rules as papers. Explicitly direct the audience to the key attribute of the visual you wish to highlight because there is no time for them to search for this visual point on each slide. Furthermore, if you choose to let the audience search on some slides, limit the number of slides that require more than cursory processing to one. For instance, use a single, relatively more complex visual slide to present the key figure showing the major quantitative finding of the study. A planned pause from rapid speaking is a powerful technique for the audience to catch their collective breath and also absorb this slide. Expecting an audience to do this 20 times in short order is unreasonable, and they will tune out. Use a separate slide to state the significance or interpretation for this finding and then begin speaking anew.

## Rule 5: Develop a consistent theme

In style, graphical design, language, and imagery, be consistent. This will ensure that the audience can allocate processing and scanning time on each slide to the salient elements that change and not to those that do not explain, support the science, or advance the main purpose of the presentation. The "branding" of your presentation and scientific message is important [3]. Use this consistency to reinforce the importance of your brand (and thus indirectly your message). Do not develop your brand using canned templates. These templates can be attractive but generally do not support the specifics of your talk and are often superfluous decoration.

## Rule 6: Repeat critical messages twice using different visuals

It is very easy to miss the main message in a rapid-fire talk, even more so than in a more traditional presentation. A total of 15 to 20 seconds to summarize the major implication or finding in a single slide is very short. Consider using a visual analog, metaphor, or simpler restatement of the major finding/implication in a subsequent slide. Typically, the assumption in these formats is that you do not cut and paste the exact same slide twice to provide oneself with more time (i.e., cheating), but you can certainly use a new slide to re-emphasize or extend the major finding. Three is a crowd and feels unduly repetitive in brief presentations. Stick with only one repetition.

## Rule 7: Use the principle of parsimony in explanations

Exactitude is as virtuous in literature as in science [13]. Identify concepts that require explanation and those that do not. Then, use simple explanations. Ensure the process or finding genuinely requires that explanation. Showing a finding and limiting what you say (Rule 4 for visuals in particular) can be a powerful means to emphasize importance. This technique also has the added benefit of providing the audience with the "space" to think, even momentarily, without distraction from the ongoing speaker dialogue. Some processes and patterns require little to no explanation. Use exactly that much. Statistics, field sampling, experimental design, and implementation strategy for the process proposed should be described in at most two succinct sentences within a 15- to 20-second interval. Explain what you need and consider engagement through less, not more, on some slides within the presentation deck.

## Rule 8: Allocate more than one slide to effectively end the narrative

At slides 16–17 in a 20-slide deck, begin closing the larger (and singular) story arc. Abrupt termination of a talk can be an effective means to jar or shock the audience but should be used sparingly—if ever. This technique comes at the cost of potential acceptance and reconciliation with the methods and implications offered. Do not leave the audience hanging. It is also natural for the audience to match the pacing and tempo of the speaker cognitively, and an abrupt end unnecessarily signals the end of a discussion and dialog.

## Rule 9: Use the final slide for contact information and links to additional resources

The total presentation time is likely a third, or less, relative to most traditional oral papers at scientific conferences. Furthermore, many rapid-fire series do not provide time for questions or feedback at the end of each presentation. This slide should reference your social media accounts, email, and website. Leverage your broader corpus of work and ideas through these links and provide a point of contact for questions. Another trick of the trade is to publish the slide deck online and provide a link to the deck within the deck at the end of the presentation. The audience will thus have an opportunity to follow-up and review the slides at a more leisurely pace if they are so inclined. Acknowledge key support, inspirations, and collaborators.

## Rule 10: Use timed practice

Speaking rapidly and clearly is not necessarily a given, even for accomplished speakers. The advancement of the slides without the speaker is a necessary condition for many of these rapid formats. Practice with the timing set in your preferred application (i.e., with autoadvance enabled via transitions between slides). There is a goldilocks effect in the number of words spoken for these formats. Too little can be awkwardly disconcerting. Too much is always disastrous. Furthermore, each slide need not suffer from the same limitations. Some require more, others less (see Rule 7). Use these differences to your advantage, and the optimal extent of description per slide can only be discovered through timed practice. Effective practice should include many of the following general approaches: stand up, speak out loud, rehearse several times without text or notes, invite an audience, record it, experiment with planned pauses, and vary pace to account for nerves or delays on the actual day. For rapid-fire talks, another common strategy is to practice with a few less seconds allocated per slide to compensate for lags when projected, audience reactions, or your movement on the stage.

## Comments

Rules are meant to broken, but not all of them and not all at once. If you elect to violate some of the rules above (best treated as suggestions), you can captivate with a story, change tempo by saying less more slowly on some slides and more on others to convey urgency, and highlight complexity without overwhelming. The audience is also an important consideration in how strictly one should consider adhering to these or any other set of proposed simple rules for scientific communication. Public talks should emphasize implications and effectively end the narrative as proposed above, whilst presentations for a group of scientists can typically invoke parsimony for explanations more directly and use appropriately technical language. The simplicity and accessibility of visuals can also be tempered by audience, and in some instances, visuals can be used to provide an analogy versus providing direct evidence or data visualization. The goal of these specific talk formats is to synthesize a topic for all audiences without a major commitment of their time. If your topic and use/misuse of the above rules stimulates some discovery for your audience and they elect to pursue the topic in greater depth, then you have absolutely succeeded. An alternative goal in considering these simple rules and in using a brief format to communicate science is to promptly share your passion for your science. If nothing else, address the "why" of the science at hand and emphasize that science is always a celebration of process and discovery. Time is up!

## References


1. Fischhoff B. The sciences of science communication. Proceedings of the National Academy of Sciences. 2013; 110:14033–9.
2. Jucan MS, Jucan CN. The Power of Science Communication. Procedia—Social and Behavioral Sciences. 2014; 149:461–6.
3. Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. Ten Simple Rules for Effective Online Outreach. PLoS Comput Biol. 2015; 11(4):e1003906. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439
4. Moser SC. Reflections on climate change communication research and practice in the second decade of the 21st century: what more is there to say? Wiley Interdisciplinary Reviews: Climate Change. 2016; 7(3):345–69.
5. Baron N. Escape from the Ivory Tower: A Guide to Making Your Science Matter. Washington, DC: Island Press; 2010.
6. Bowater L, Yeoman K. Science Communication—A Practical Guide for Scientists. Oxford, UK: Wiley-Blackwell; 2013.
7. Bickford D, Posa MRC, Qie L, Campos-Arceiz A, Kudavidanage EP. Science communication for biodiversity conservation. Biological Conservation. 2012; 151(1):74–6. http://dx.doi.org/10.1016/j.biocon.2011.12.016.
8. Bultitude K. The Why and How of Science Communication. In: Rosulek P, editor. Science Communication. Europe: Pilsen: European Commission; 2011.
9. Greenhow C, Gleason B. Social scholarship: Reconsidering scholarly practices in the age of social media. British Journal of Educational Technology. 2014; 45(3):392–402.
10. Van Noorden R. Online collaboration: Scientists and the social network. Nature. 2014; 512:126–9. https://doi.org/10.1038/512126a PMID: 25119221
11. McClain C, Neeley L. A critical evaluation of science outreach via social media: its role and impact on scientists. F1000 Research. 2015; 3:300.
12. Valkenburg PM, Peter J, Walther JB. Media Effects: Theory and Research. Annual Review of Psychology. 2016; 67(1):315–38.

**13.** Calvino I. Six Memos for the Next Millennium. Cambridge, Massachusetts: Harvard University Press; 1988.

**14.** Bourne PE. Ten Simple Rules for Making Good Oral Presentations. PLoS Comput Biol. 2007; 3(4):e77. https://doi.org/10.1371/journal.pcbi.0030077 PMID: 17500596

**15.** Rougier NP, Droettboom M, Bourne PE. Ten Simple Rules for Better Figures. PLoS Comput Biol. 2014; 10(9):e1003833. https://doi.org/10.1371/journal.pcbi.1003833 PMID: 25210732

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Writing a Reply Paper

**Mark P. Simmons***

Department of Biology, Colorado State University, Fort Collins, Colorado, United States of America

* psimmons@lamar.colostate.edu

## Introduction

> *Efforts to subject one's own pet hypotheses to severe tests, to attempt to falsify them, are always warranted. The likelihood that one will impose unrealistically high standards is quite low. The same cannot be said for the standards that scientists set for the work of their opponents. . . . As unseemly as factionalism in science may be, it does serve a positive function. It enlists baser human motives for higher causes.—David L. Hull*, Science as a Process: An Evolutionary Account of Social and Conceptual Development of Science [1]

Have you read a paper or watched a presentation at a scientific conference with an utterly wrong or unsubstantiated conclusion? Have you participated in a discussion group wherein the group identified a fatal flaw in a recently published paper? The simple thing to do is discard the paper that you just read, ask a tough question at the end of the presentation, and make jokes about the paper with your colleagues at the end of the discussion, respectively. By the time students complete their PhD, they have probably seen instances of all three of these. But rather than stop there, it may be appropriate to invest substantial extra effort to write a formal reply.

## Rule 1: Determine If It Is Appropriate to Write a Reply

If your focus is on one or two papers, then a reply is the proper forum. However, if you want to address numerous related papers and advocate a shift towards better practices for an entire discipline, then you may be better off writing a perspectives paper for the applicable *Annual Reviews* or *Trends* journals (though these generally require invitations).

Writing replies is not a research program in and of itself, and you should not expect to make a successful career out of writing replies. You cannot get grants for writing replies, but you can make plenty of adversaries who go on to serve as your funding-agency program officers or panelists, editors, and reviewers. In the same way that you should carefully pick your fights, you should exercise discretion about when to write a reply.

If the original paper is scientifically uninteresting, then your reply will probably be equally uninteresting and not worth writing. One exception to this is when the original paper was published in a prominent journal. A classic example is Williamson's [2] *Proceedings of the National Academy of Sciences of the USA* (PNAS) paper, wherein he asserted that larvae and adults did not evolve from a single common ancestor. This fantastic communicated submission from an NAS member was immediately rebutted by a direct submission from Hart and Grosberg [3]. That same year the editor in chief of PNAS [4] announced that the journal would stop accepting communicated submissions.

CrossMark
click for updates

OPEN ACCESS

**Citation:** Simmons MP (2015) Ten Simple Rules for Writing a Reply Paper. PLoS Comput Biol 11(10): e1004536. doi:10.1371/journal.pcbi.1004536

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** October 8, 2015



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

## Rule 2: Do Not Be Intimidated by Big Data—or Big Names

In this Age of Omics [5] (e.g., genomics, proteomics, and transcriptomics) and massive publicly available data repositories (e.g., Dryad, GenBank, and The Arabidopsis Information Resource), you should not assume that the original authors' biologically implausible yet impressively well-supported conclusions are correct. Big data have the potential to create impressive *p*-values for incorrect conclusions even when there is a small but consistent underlying bias. Find the underlying bias, and the impressive *p*-values may vanish. In this context, understanding the limitations of the analytical methods applied by the original authors and knowing what patterns to look for while examining their data are more important than being able to feed the data through bioinformatic pipelines. Too many authors rely upon summary statistics generated by pipelines.

Early in this Age of Omics, the International Human Genome Sequencing Consortium [6] claimed to have identified 223 genes that had been laterally transferred from bacteria to vertebrate nuclear genomes based on Protein BLAST scores. By contrast, Salzberg et al. [7] asserted that the earlier claim could be more plausibly explained by rate heterogeneity among lineages together with taxonomic undersampling and gene loss in nonvertebrate eukaryotic genomes. Salzberg et al.'s [7] successful rebuttal was based on their understanding the limits of BLAST searches, making appropriate null hypotheses, and examining the original data in a phylogenetic context; they relied on first principles rather than attempting to develop new algorithms or sequence new genomes.

In the same manner that you should not be intimidated by big data, you should not be intimidated by big names either. This is not a new idea—it's widespread on buttons and bumper stickers: "Question Authority." Wenzel [8] provided the following fun example from his undergraduate studies that I tell all of my students:

> For example, when I was studying thermodynamics in introductory physics in 1977, my professor (a Nobel Laureate) explained how the rate of heat production and heat loss limits mammals to be no smaller than a shrew, and of course, shrews are the smallest mammals. A student in the front row asked "Dr. Purcell, what about baby shrews?" The professor's jaw dropped; he had never thought of that.

## Rule 3: Stick to the Facts

Do not get personal by writing inflammatory statements. Instead, let the evidence speak for itself—you are writing for an intelligent, specialist audience that is capable of reaching its own conclusions. Furthermore, do not write in a righteous manner, because there is always the possibility that you are wrong. Your reply will become part of the permanent scientific record, and a cute but false remark might still be talked about by your colleagues 20 years hence.

You may have fond memories of reading flamboyant reply papers (e.g., [9,10]). But unless you are at the top of your field and have a close relationship with the editor, do not expect to get away with it because you will probably be forced to remove the flamboyant text by the reviewers and editors. Even if you are at the top of your field, you still have to live with the long-term consequences of upsetting the original authors. For example, as of 2015, Sudhir Kumar is the editor in chief of *Molecular Biology and Evolution*, while William Martin is the editor in chief of its sister journal, *Genome Biology and Evolution*. These two editors are expected to coordinate these journals, irrespective of their aggressive replies to each other [9,11].

## Rule 4: Expect a Reply to Your Reply

Your reply will be held to a very high standard by the original authors, and rightfully so. Expect to be quoted, with "[sic]" included, where applicable, so carefully proofread your manuscript. Scientific papers should always be readable literally. However, that fact is especially relevant to how the original authors may reply to you. If you overstate your case by making absolute assertions, then expect their reply to focus on that. You are the one picking a fight, so you better win it.

## Rule 5: Make Points of General Interest

Try to make points of general interest rather than just writing for the small target audience who read the original paper. This approach is relevant for the sake of getting your reply accepted into a prominent journal (ideally the same journal in which the original paper was published) and having your reply cited. If all you do in your reply is successfully refute the original paper, then you have made an important scientific contribution, but few people will ever cite your reply (or the original paper) because the case is closed.

Having your reply published in the same journal as the original paper helps you reach the readers of the original paper. If these readers do not read your reply, then they may think that the applicable conclusions from the original paper are still valid. In addition to your choice of journal, another way to increase the prominence of your reply is to make it a community effort by inviting colleagues to join your reply as coauthors. For example, Melissa Luckow [12] demonstrated wide support for her viewpoint that the type species of the legume genus *Acacia* should not be moved from Africa to Australia by teaming up with 36 other botanists whose work would be affected by the taxonomic change proposed by Orchard and Maslin [13].

If you develop a reputation for writing informative reply papers, then many of your colleagues will read your reply even if they have not read the original paper so that they can better understand the issues being discussed. Help your readers out by objectively summarizing the original paper, providing the necessary context, and, if applicable, clarifying the assertion(s) made by the original authors in your introduction.

## Rule 6: Be Positive Too

Rather than just pointing out flaws, present an alternative analytical method or empirical conclusion that improves upon the method or conclusion from the original paper. As Theodore Roosevelt famously stated in his 1910 speech at the Sorbonne [14], "It is not the critic who counts; not the man who points out how the strong man stumbles, or where the doer of deeds could have done them better. The credit belongs to the man who is actually in the arena. . .."

There are many classic replies for which the original papers have been largely forgotten but the reply lives on because the authors of the reply presented an important novel alternative. Nixon [15] demonstrated the efficiency of his novel Parsimony Ratchet by showing that it found shorter phylogenetic trees with dramatically less computing power (a desktop in less than two hours versus Sun workstations in months) than Rice et al.'s [16] traditional tree search on a 500-taxon dataset. Nixon [15] rendered a direct reply to Rice et al. [16] unnecessary because he showed that their tree-search strategy was obsolete. According to Web of Science, Nixon [15] was cited 86 times in 2014, whereas Rice et al. [16] was not cited at all.

## Rule 7: Only Present Your Best Arguments

Do not present all possible arguments. If you do so, then a reply by the original authors might only address your weakest arguments, yet casual readers will think that your reply has been invalidated because the majority of your arguments have been successfully rebutted or at least muddled. Instead focus on your most important arguments and clearly itemize them so they cannot be overlooked by the original authors should they choose to write their own reply. You do not have to refute every point made in the original paper to write an effective reply. It is enough to refute their primary conclusion(s) rather than pedantically going after every point.

## Rule 8: Use the Authors' Own Arguments against Them

If applicable, demonstrate that the original authors' conclusion is falsified by their own criteria rather than just evaluating their conclusion using your own criteria. James S. Farris has published numerous such replies. In perhaps his most important reply, Farris [17] demonstrated that most parsimonious trees, rather than trees constructed using overall similarity, have the highest information content. By doing so, Farris demonstrated that the primary goal of phenetics [18] was best solved using cladistic methods, thereby effectively undermining the justification for using phenetics for taxonomic classifications.

## Rule 9: Present the Logical End Point of Faulty Arguments

If applicable, demonstrate that the original authors' argument is untenable when it is taken to its logical conclusion. For example, Mollison [19] demonstrated that the Lotka-Volterra equations, when applied to predators and their prey [20], can require that the prey population recover from a tiny fraction ($10^{-18}$) of an individual, which Mollison famously referred to as an "atto-fox" (atto being the prefix for $10^{-18}$ and fox being the [rabies virus'] prey in his empirical example).

## Rule 10: Demonstrate That Conclusions Can Be Explained by the Null Hypothesis

If the original authors' assertions are untestable, tautological, or can be equally well explained by the null hypothesis, then they are uninformative. Connor and Simberloff [21] demonstrated that one or more of these flaws apply to all seven of Diamond's [22] ecological assembly rules, thereby refuting the bases for Diamond's inductive conclusion that interspecific competitive exclusion is a primary factor in determining bird communities.

## Acknowledgments

The author thanks Kevin Nixon for teaching him rule 7; Dan Sloan for several discussions; and Jim Carpenter, Jerry Davis, Jeff Doyle, John Freudenstein, John Gatesy, Kim Hoke, Melissa Luckow, Kevin Nixon, Vern Simmons, and John Wenzel for numerous helpful suggestions on earlier drafts of this manuscript.

## References

1. Hull DL. Science as a process: an evolutionary account of social and conceptual development of science. Chicago: The University of Chicago Press; 1988.
2. Williamson DI. Caterpillars evolved from onychophorans by hybridogenesis. P Natl Acad Sci USA. 2009; 106: 19901–19905.
3. Hart MW, Grosberg RK. Caterpillars did not evolve from onychophorans by hybridogenesis. P Natl Acad Sci USA. 2009; 106: 19906–19909.
4. Schekman R. PNAS will eliminate communicated submissions in July 2010. P Natl Acad Sci USA. 2009; 106: 15518.
5. Abu-Issa R, Kirby ML. Take heart in the age of "omics." Circ Res. 2004; 95: 335–336. PMID: 15321941

6. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860–921. PMID: 11237011

7. Salzberg SL, White O, Peterson J, Eisen JA. Microbial genes in the human genome: lateral transfer or gene loss? Science 2001; 292: 1903–1906. PMID: 11358996

8. Wenzel JW. Why I teach introductory courses. In: Talking about teaching: essays by members of The Ohio State University Academy of Teaching. Columbus: The Ohio State University Academy of Teaching; 2004. vol. 1, pp. 23–32.

9. Graur D, Martin W. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 2004; 20: 80–86. PMID: 14746989

10. Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol. 2013; 5: 578–590. doi: 10.1093/gbe/evt028 PMID: 23431001

11. Hedges SB, Kumar S. Precision of molecular time estimates. Trends Genet. 2004; 20:242–247. PMID: 15109778

12. Luckow M, Hughes C, Schrire B, Winter P, Fagg C, Fortunato, et al. *Acacia*: the case against moving the type to Australia. Taxon 2005; 54: 513–519.

13. Orchard AE, Maslin BR. (1584) Proposal to conserve the name *Acacia* (*Leguminosae*: *Mimosoideae*) with a conserved type. Taxon 2003; 52: 362–363.

14. Almanac of Theodore Roosevelt. Citizenship in a republic; 2015. http://www.theodore-roosevelt.com/images/research/speeches/maninthearena.pdf. Accessed 14 April 2015.

15. Nixon KC. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 1999; 15: 407–414.

16. Rice KA, Donoghue MJ, Olmstead RG. Analyzing large data sets: *rbcL* 500 revisited. Syst Biol. 1997; 46: 554–563. PMID: 11975333

17. Farris JS. The information content of the phylogenetic system. Syst Zool. 1979; 28: 483–519.

18. Sokal RR, Sneath PHA. Principles of numerical taxonomy; 1963. San Francisco: W.H. Freeman and Co.

19. Mollison D. Dependence of epidemic and population velocities on basic parameters. Math Biosci. 1991; 107: 255–287. PMID: 1806118

20. Lotka AJ. Elements of physical biology. Baltimore: Williams & Wilkins Company; 1925.

21. Connor EF, Simberloff D. The assembly of species communities: chance or competition? Ecology 1979; 60: 1132–1140.

22. Diamond JM. Assembly of species communities. In: Cody ML, Diamond JM, editors. Ecology and evolution of communities. Cambridge: Belknap Press; 1975. pp. 342–444.

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple (Empirical) Rules for Writing Science

**Cody J. Weinberger**[1], **James A. Evans**[2,3], **Stefano Allesina**[1,3]*

1 Department of Ecology & Evolution, University of Chicago, Chicago, Illinois, United States of America,
2 Department of Sociology, University of Chicago, Chicago, Illinois, United States of America,
3 Computation Institute, University of Chicago, Chicago, Illinois, United States of America

* sallesina@uchicago.edu

> *". . .though a Philosopher need not be sollicitous that his style should delight its Reader with his Floridnesse, yet I think he may very well be allow'd to take a Care that it disgust not his Reader by its Flatness, especially when he does not so much deliver Experiments or explicate them, as make Reflections or Discourses on them; for on such Occasions he may be allow'd the liberty of recreating his Reader and himself, and manifesting that he declin'd the Ornaments of Language, not out of Necessity, but Discretion. . ."—Robert Boyle, Proëmial Essay* [1].

Scientists receive (and offer) much advice on how to write an effective paper that their colleagues will read, cite, and celebrate [2–15]. Fundamentally, the advice is similar to that given to journalists: keep the text short, simple, bold, and easy to understand. Many resources recommend the parsimonious use of adjectives and adverbs, the use of present tense, and a consistent style. Here we put this advice to the test, and measure the impact of certain features of academic writing on success, as proxied by citations.

The abstract epitomizes the scientific writing style, and many journals force their authors to follow a formula—including a very strict word-limit, a specific organization into paragraphs, and even the articulation of particular sentences and claims (e.g., "Here we show that. . .").

For our analysis, we collected more than one million abstracts from eight disciplines, spanning 17 years. The disciplines were chosen so that biology was represented by three allied fields (Ecology, Evolution, and Genetics). We drew upon a wide range of comparison disciplines, namely Analytic Chemistry, Condensed Matter Physics, Geology, Mathematics, and Psychology (see table in S1 Text). We measured whether certain features of the abstract consistently led to more (or fewer) citations than expected, after accounting for other factors that certainly influence citations, such as article age (S1 Fig), number of authors and references, and the journal in which it was published.

We organized the most frequent suggestions into "Ten Simple Rules," and probed them by testing a variety of features from the abstracts. Because the style and requirements for abstracts can vary dramatically between journals (S2 Fig), we normalized all the measures according to their distribution for each journal (S1 Text).

## Rule 1: Keep It Short

This is the most universally accepted piece of advice given to writers [3,7,9,11–13]. We tested this by examining the effect of shorter abstracts on citation, measuring the number of words (Rule 1a [R1a]) and number of sentences (R1b) in each abstract.

CrossMark
click for updates

OPEN ACCESS

**Citation:** Weinberger CJ, Evans JA, Allesina S (2015) Ten Simple (Empirical) Rules for Writing Science. PLoS Comput Biol 11(4): e1004205. doi:10.1371/journal.pcbi.1004205

**Published:** April 30, 2015



**Funding:** CJW and SA are supported by NSF #1042164, JAE by NSF #1158803. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Rule 2: Keep It Compact

The typical advice is to keep sentences or phrasing short, break compound sentences into simpler sentences, and remove any "unnecessary" words [2–6,9–12,14]. We evaluated this by measuring the effect of having sentences shorter than the mean for the journal where the article was published (R2).

## Rule 3: Keep It Simple

Canonical advice includes the prescription to use plain language and avoid jargon and technical terms [2–4,7,10,12,14]. Many of the most prominent journals state that their abstracts should be accessible to scientists working in different disciplines. To test this, we measured the proportion of words in the abstract that are found in a standard English dictionary (R3a) and that are present in a dictionary of "easy words" (R3b).

## Rule 4: Use the Present Tense

Stylists recommend the use of the present tense [10,12], as it is more direct and deemed easier to understand for non-native speakers. We assessed this by ascertaining the ratio of (present tense)/(present + past tense) (R4).

## Rule 5: Avoid Adjectives and Adverbs

Using few adjectives and adverbs avoids fluff and keeps the text short and easy to understand [4,8,9,12]. We measured the effect of having a proportion of adjectives and adverbs smaller than that typical for the journal (R5).

## Rule 6: Focus

Many authors suggest sticking to a single point, and reiterating the "take home" message [5,6,11,13,14]. We captured this with the proportion of words in the abstract that were also keywords (R6).

## Rule 7: Signal Novelty and Importance

There is conflicting advice on whether to explicitly state the significance of your work. Stressing that the work is novel and solves important problems helps to "sell" the article [12,15]. Opponents of this rule say that all published work should already meet these criteria [8,13]. We examined this by checking whether the abstract contained at least one word signaling novelty (e.g., "novel," "new," "innovative" [R7a]) and, separately, a word signaling importance (e.g., "key," "significant," "crucial" [R7b]).

## Rule 8: Be Bold

Many authors suggest "selling" the work forcefully and stressing positive results. We tested this by measuring the ratio (superlatives)/(superlatives + comparatives) (R8).

## Rule 9: Show Confidence

Similarly, using too many "hedge words" (e.g., "somewhat," "speculative," "appear," "almost," "largely") can signal a lack of confidence in the work. We explored this with the measure of fewer hedge words in the abstract (R9).

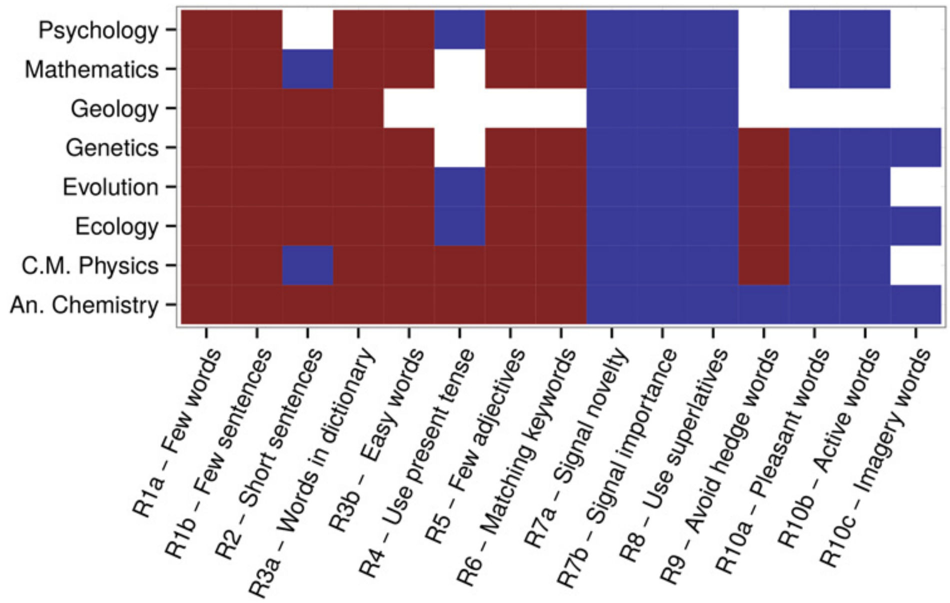



**Fig 1. Effect of abstract features on citations.** For each discipline (rows) and each abstract feature (columns), we measured whether a certain feature (e.g., having fewer words than the typical abstract published in the same journal [R1a]) led to a significant increase (blue) or decrease (red) in total citations. We considered an effect positive or negative only if the associated probability of being zero was smaller than 0.01/15 (i.e., we applied the Bonferroni correction to obtain an overall significance level of 1%).

doi:10.1371/journal.pcbi.1004205.g001

## Rule 10: Avoid Evocative Words

A style perceived as too flowery or involving the overuse of highly evocative words is discouraged. We tested whether using words perceived as "pleasant," "active," or "easy to imagine" led to more citations than those for abstract containing "unpleasant," "passive," or "hard to imagine" words [16–18] (R10a–c).

## Results

In Fig 1, we report the sign of the effect associated with each abstract feature (column) for each discipline (row). Surprisingly, half of the typical suggestions—including those that are most common, about brevity and clarity—are associated with a significant decrease in citations.

We find that shorter abstracts (fewer words [R1a] and fewer sentences [R1b]) consistently lead to fewer citations, with short sentences (R2) being beneficial only in Mathematics and Physics. Similarly, using more (rather than fewer) adjectives and adverbs is beneficial (R5). Also, writing an abstract with fewer common (R3a) or easy (R3b) words results in more citations.

The use of the present tense (R4) is beneficial in Biology and Psychology, while it has a negative impact in Chemistry and Physics, possibly reflecting differences in disciplinary culture.

While matching the keywords (R6) leads to universally negative outcomes, signaling the novelty and importance of the work (R7) has positive effects. The use of superlatives (R8) is also positive, while avoiding "hedge" words is negative in Biology and Physics, but positive in Chemistry.

Finally, choosing "pleasant," "active," and "easy to imagine" words (R10) has positive effects across the board.

When we measured effect sizes (Fig 2), we found that abstract features can have a strong influence on citations. Being one standard deviation above the mean for a given feature (with respect to the mean for corresponding journal) can increase citations by 4.6% (Mathematics

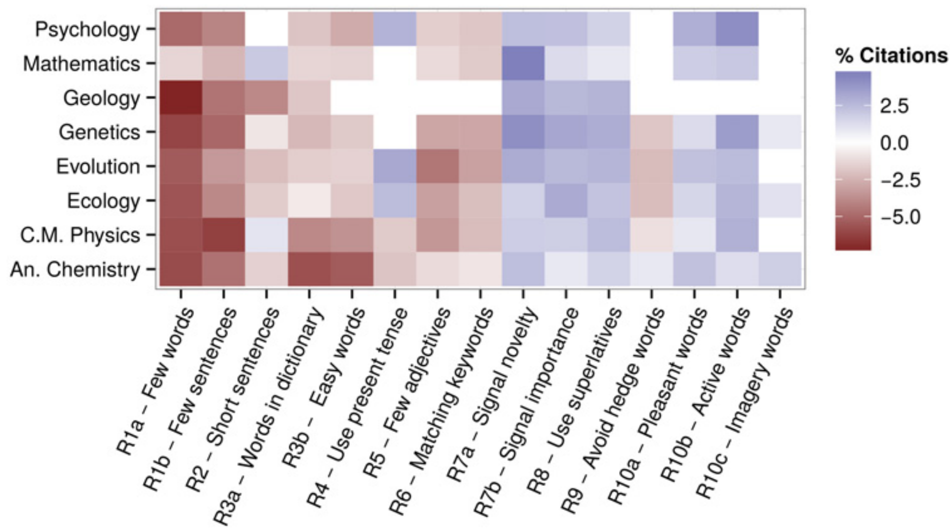



**Fig 2. Size of the effects.** Same designations as Fig 1, but measuring the benefit/cost of having a certain feature one standard deviation above the mean for the corresponding journal.

doi:10.1371/journal.pcbi.1004205.g002

[R7a]), or decrease them by 7.2% (Geology [R1a]). When analyzing each journal separately, we find qualitatively the same results (S3–S10 Figs).

## Conclusions

We have found that—when it comes to abstracts—"more is more," despite clear and abundant advice to the contrary.

This is an interesting and surprising result. An intriguing hypothesis is that scientists have different preferences for what they would like to read versus what they are going to cite. Despite the fact that anybody in their right mind would prefer to read short, simple, and well-written prose with few abstruse terms, when building an argument and writing a paper, the limiting step is the ability to find the right article. For this, scientists rely heavily on search techniques, especially search engines, where longer and more specific abstracts are favored. Longer, more detailed, prolix prose is simply more available for search. This likely explains our results, and suggests the new landscape of linguistic fitness in 21st century science. Future studies could investigate the relationship between stylistic features and retrievability directly, as well as the strength of the relationship between retrievability and citation performance.

Another interesting finding is that there is very little variation across disciplines, with only three out of fifteen features displaying sign changes among the diverse fields we examined.

Scientists are skeptical by disposition, and this exercise shows that, rather than taking advice at face value, they can apply the same machinery they use to interrogate nature to put these recommendations to the test—and write a lengthy, convoluted, highly-indexible, self-describing abstract.

## Supporting Information

**S1 Text. Supporting Methods and Results.** Description of the data, the features analyzed and the statistical models; discipline-specific results.
(PDF)

**S1 Fig. Distribution of citations through time.** Figure showing that citations received by the articles in a journal/year combination are approximately log-normally distributed.
(TIFF)

**S2 Fig. Number of words in abstracts.** Distribution of the number of words in the abstract divided by discipline.
(TIFF)

**S3 Fig. Effect sizes in Analytical Chemistry.** As Fig 2, but analyzing Analytical Chemistry journals.
(TIFF)

**S4 Fig. Effect sizes in Ecology.** As Fig 2, but analyzing Ecology journals.
(TIFF)

**S5 Fig. Effect sizes in Evolution.** As Fig 2, but analyzing Evolution journals.
(TIFF)

**S6 Fig. Effect sizes in Genetics.** As Fig 2, but analyzing Genetics journals.
(TIFF)

**S7 Fig. Effect sizes in Geology.** As Fig 2, but analyzing Geology journals.
(TIFF)

**S8 Fig. Effect sizes in Mathematics.** As Fig 2, but analyzing Mathematics journals.
(TIFF)

**S9 Fig. Effect sizes in Condensed Matter Physics.** As Fig 2, but analyzing Condensed Matter Physics journals.
(TIFF)

**S10 Fig. Effect sizes in Psychology.** As Fig 2, but analyzing Psychology journals.
(TIFF)

## Acknowledgments

Thanks to G. Barabás, M. Begun, J. Grilli, P. McMahan, E. Sander, M.J. Smith, and M. Teplitskiy for comments and discussion.

## References

1. Boyle R. A proemial essay, wherein, with some considerations touching experimental essays in general, is interwoven such an introduction to all those written by the author, as it is necessary to be perus'd for the better understanding of them. In: Certain physiological essays and other tracts written at distant times, and on several occasions by the honourable Robert Boyle; wherein some of the tracts are enlarged by experiments and the work is increased by the addition of a discourse about the absolute rest in bodies. 2nd ed. Henry Herringman, republished by University of Michigan, Digital Library Production Service; 1669. p. 12–13. Available from: http://quod.lib.umich.edu/e/eebo/A28944.0001.001?view = toc.
2. Paul JK. Scientific writing. Oral Surgery, Oral Medicine, Oral Pathology. 1970; 30(2):185–191. PMID: 5270920
3. Lilleyman J. How to write a scientific paper–a rough guide to getting published. Archives of disease in childhood. 1995; 72(3):268. PMID: 7741582
4. Evans M. Writing for publication. British Journal of Oral and Maxillofacial Surgery. 1998; 36(3):161–164. PMID: 9786649
5. Alexandrov AV. How to write a research paper. Cerebrovascular diseases. 2004; 18(2):135–138. PMID: 15218279

6. Chiswick M. Writing a research paper. Current Paediatrics. 2004; 14(6):513–518.
7. Cunningham S. How to... write a paper. Journal of Orthodontics. 2004; 31(1):47–51. PMID: 15071152
8. Thrower PA. Writing a scientific paper: I. Titles and abstracts. Carbon. 2007; 45(11):2143–2144.
9. Van Way CW. Writing a scientific paper. Nutrition in Clinical Practice. 2007; 22(6):636–640. PMID: 18042951
10. Fahy K. Writing for publication: the basics. Women and Birth. 2008; 21(2):86–91. doi: 10.1016/j.wombi.2007.12.005 PMID: 18282754
11. Christensen NB, Kume H, Autorino R. How to write titles and abstracts for readers. International Journal of Urology. 2009; 16(1):2–3. doi: 10.1111/j.1442-2042.2008.02228.x PMID: 19120521
12. Davidson A, Delbridge E. How to write a research paper. Paediatrics and Child Health. 2012; 22(2):61–65.
13. Mack C. How to write a good scientific paper: title, abstract, and keywords. Journal of Micro-Nanolithography MEMS and MOEMS. 2012; 11(2):020101.
14. Cals JW, Kotz D. Effective writing and publishing scientific papers, part II: title and abstract. Journal of clinical epidemiology. 2013; 66:585. doi: 10.1016/j.jclinepi.2013.01.005 PMID: 23434329
15. Reis SRN, Reis AI. How to write your first scientific paper. In: Interdisciplinary Engineering Design Education Conference (IEDEC), 2013 3rd. IEEE; 2013. p. 181–186.
16. Sweeney K, Whissell C. A dictionary of affect in language: I. Establishment and preliminary validation. Perceptual and motor skills. 1984; 59(3):695–698.
17. Whissell C. The dictionary of affect in language. Emotion: Theory, research, and experience. 1989; 4(113–131):94.
18. Whissell C. Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language 1, 2. Psychological reports. 2009; 105(2):509–521. PMID: 19928612

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Effective Online Outreach

**Holly M. Bik[1,2]*, Alistair D. M. Dove[3], Miriam C. Goldstein[4], Rebecca R. Helm[5], Rick MacPherson[6], Kim Martini[7,8], Alexandria Warneke[9], Craig McClain[10]**

**1** UC Davis Genome Center, University of California—Davis, Davis, California, United States of America, **2** School of Biosciences, University of Birmingham, Birmingham, United Kingdom, **3** Georgia Aquarium Research Center, Atlanta, Georgia, United States of America, **4** California Sea Grant, La Jolla, California, United States of America, **5** Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island, United States of America, **6** Pelagia Consulting, San Francisco, California, United States of America, **7** Joint Institute for the Study of Atmosphere and Ocean, University of Washington, Seattle, Washington, United States of America, **8** NOAA/Pacific Marine Environmental Laboratory, Seattle, Washington, United States of America, **9** Department of Biology, San Diego State University, San Diego, California, United States of America, **10** National Evolutionary Synthesis Center, Durham, North Carolina, United States of America

* holly.bik@gmail.com



OPEN ACCESS

**Citation:** Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. (2015) Ten Simple Rules for Effective Online Outreach. PLoS Comput Biol 11(4): e1003906. doi:10.1371/journal.pcbi.1003906

**Editor:** Philip E. Bourne, National Institutes of Health, United States of America

**Published:** April 16, 2015



**Funding:** HMB acknowledges funding from the Alfred P. Sloan Foundation, the North Pacific Research Board, and the National Science Foundation (DBI-1262480). KM acknowledges Pacific Marine Environmental Laboratory Contribution Number 4174 and Joint Institute for the Study of the Atmosphere and Ocean Contribution Number 2295. The funders had no role in the preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

Online science outreach is paradoxically both easy and difficult. While anyone can start a blog and post updates to Twitter, it can be extremely challenging to establish a long-term following and demonstrate solid measures of success. A daunting number of online tools and platforms exist, and choosing where to start can be a difficult task in itself (for an explanation and guide to online tools, see [1]). As practicing scientists who have contributed to the highly visited marine science blog Deep-Sea News (DSN) for up to nine years, we provide guidance on how scientists, who often have minimal excess time and more pressing priorities, can maximally utilize new media tools. Here, we describe ten rules for conducting effective online outreach, so that other scientists can also enjoy the advantages of disseminating their knowledge and expertise through social media.

## Background and Overview of Deep-Sea News

Deep-Sea News (http://deepseanews.com) was established in 2005 with the sole purpose of tracking new literature for deep-sea scientists. Over the years, DSN content gradually transformed into a blog, and then incorporated other social media tools to become a multichannel online outreach platform. Currently, DSN is written by eight scientific professionals (the authors of this manuscript, with expertise in genomics, computational biology, ecology, oceanography, and conservation), with a collective expertise across a wide range of marine and oceanographic sciences. Capitalizing on our diverse interests and various social outlets, our readership has expanded to make DSN one of the most popular marine science blogs on the Internet, with ~7,000 average hits/day, and more than 8,000,000 cumulative hits. Because of our high visibility, DSN authors are routinely cited by various science journalists and quoted in prominent general audience publications such as Slate, Wired, National Geographic, and the annual print anthology The Open Laboratory: The Best Science Writing Online. We believe that the success of Deep-Sea News as an effective tool for online outreach is best illustrated by our mission statement: our commitment to "demystify and humanize science in an open conversation that instills passion, awe, and responsibility for the oceans" (http://deepseanews.com/about-2/mission/).

## Rule 1: Stop Treating Outreach and Research As Separate Entities

Integrating outreach tightly with ongoing research projects is key for maximizing success and minimizing time investment. Consider writing blog posts about either your newly published paper, literature you have compiled for manuscripts or grant writing, or literature you are reading for journal clubs. Blog posts focused on papers you have co-authored can serve as an important long-term resource, both as a first-person press release and a reference point for journalist enquiries. One example of research-outreach integration at DSN is the "Sizing Ocean Giants" project (http://www.storyofsize.com/sizing-ocean-giants/), an undergraduate research project focused on marine organism body size that included a heavy social media component. Undergraduates were asked to use social media to engage the public about their specific animals, leveraging online platforms as a pedagogical tool to increase their understanding of their organisms and the scientific process. Social media also served as a mechanism to engage other researchers. For example, a marine mammalogist reached out to the group via Twitter and ended up contributing large datasets to the project.

Since career advancement in the sciences depends on research productivity, it is important to link traditional metrics with online work. Based on our own experience, new grant proposals, research collaborations, and manuscripts (e.g., this article) can often result from online interactions. It should also be acknowledged that the requirement of translating research to a public audience increases both awareness and intimacy with the published literature—one that can feed directly back into your research program. Researchers can benefit from engaging with a broader set of disciplines than one would normally interact with, and these combined factors can lead to origination of new research ideas.

Online outreach projects can be designed and organized in a way that equates to publishable units; for example, can the data ultimately be used as the basis of a research paper, or can the outreach experience itself represent a useful case study that could be conveyed as an editorial or commentary? Research and teaching activities should also be considered in light of the outreach potential. In the Sizing Ocean Giants project described above, students used new media tools as part of their class projects. Another example is tweeting from conferences, which can quickly broaden your professional network. Considering the above points will ensure that your outreach activities and social media use work hand-in-hand with your research program. We also note here that building a track record in online outreach is, in itself, important for justifying your ability to successfully execute Broader Impacts sections of grant proposals.

## Rule 2: Be Strategic. Be Deliberate

Before embarking on any actual online outreach, the first step should be to define your overall short-term and long-term goals. We believe that it is important to incorporate some type of formal planning mechanism for all online outreach projects; planning is key for defining, measuring, and evaluating success over the course of the project. The DSN mission statement and core values (http://deepseanews.com/about-2/mission/) resulted from a facilitated, in-person blog retreat in October 2011. DSN needed a clear vision as to where our social media outreach was headed, including our niche, our goals, and our values. We could not measure success unless we defined what success was. Furthermore, it was important to explicitly iterate these ideas on the website itself, so that our audience knew what to expect from us.

Despite the informality of our group, we took this process very seriously and ran the retreat as a mediated strategic planning session, complete with mission statements, value propositions and action items. This clarifying of our mission and core values allowed us to build more effective strategies for generating content, attracting new members, and building an audience. As a result, our website growth in 2012 (as measured by daily hit rate) changed from a linear to an exponential growth trajectory, which continues until today.

## Rule 3: Find *Your* Niche and Story

After drafting your outreach goals, the next step is to determine your target audience, define your outreach angle (i.e., a writing voice or online persona), and determine the online tools that are best suited to your needs. Don't assume that online audiences are only interested in dinosaurs, sex, or chocolate (reoccurring subjects on many mainstream science blogs)—let your own scientific passions drive the content you generate and curate (see also Rule 6). For example, Chris Mah, a researcher at the Smithsonian Institution, is unapologetic about his enthusiasm for echinoderms (http://echinoblog.blogspot.com/). Also keep in mind that any posts that go viral (or touch on controversial topics) can swing a reader's view of what a blog is all about. Because it is difficult to predict popularity, it is worthwhile to ask if a post's content is representative of how you want people to know your blog.

Here we note that outreach projects should consciously choose between "inreach" versus "outreach." Inreach refers to discussion and sharing amongst a known, closed group (most likely a community to which you already belong). Using Twitter to discuss a research technique with colleagues in your subdiscipline is one example of inreach. Other examples may include submitting code patches on GitHub, or commenting on peer-reviewed publications via PubMed commons. Although it may increase the visibility of the conversation due to the public nature of the platform, the discussion is not likely to extend beyond scientists, and certainly not to the general public. Outreach, on the other hand, refers to a dedicated and sustained effort to disseminate science beyond the ivory tower, for example, working with journalists to discuss new research in the mainstream media, or conducting an "ask me anything" live blogging session on a site such as Reddit. Your specific goals will determine whether your project is best suited to inreach, outreach, or some combination of the two. There is a place for both avenues online, but remember that these two concepts are not the same at all.

## Rule 4: Branding. . .Branding. . .Branding. . .

Branding may seem "corporate", but style and consistency are both critical for online outreach. Branding is a powerful tool. In the corporate arena, a well-established brand can be the key to a successful business enterprise. The same ideology holds true in the online ecosystem. Fundamentally, a brand represents a promise to your audience that you will abide by a certain set of principles or a mission statement. The best brands are ones that instill and inspire others to extend a personal interest and trust in the message the brand exemplifies. This brand should permeate all aspects of your platform and instill a sense of quality, credibility, and experience. However, it is important to note that such a trust does not happen quickly, and effective branding can require a significant investment of time and energy.

Before officially branding any online outreach platform, you must first solidify the message you are trying to convey and the audience for whom it is intended. At DSN, our mission statement and set of core values has set forth a clear and consistent style—a style which our readership has come to expect. This standard allows a diversity of voices to be present through multiple outlets, while keeping those voices true to our common objective. We believe our emphases on "saying things that others do not" and "awareness through scrutiny, not negativity" (http://deepseanews.com/about-2/mission/) have greatly added to our brand and successes in online outreach. Moreover, we have found that having a clean, easily accessible, and visually appealing interface is also beneficial for branding and building readership. Most importantly, we have chosen a distinctive symbol, the giant squid, that is consistent, recognizable, and personifies what DSN stands for (http://deepseanews.com/2011/06/from-the-editors-desk-the-giant-squid-can-be-a-panda-for-the-ocean/). Our logo also incorporates a homage to marine field work (red/white diagonal akin to a SCUBA flag), and a sense of playfulness (a pirate's eye-patch). Taken together, our branding encapsulates the underlying core values and mission of DSN.

## Rule 5: Recruit a Top-Notch Team

Because social media is so much work, distributing effort and delegating tasks amongst different participants can greatly increase long-term sustainability. Online outreach requires producing regular content, appealing to diverse users, and building a long-term community of readers and commenters. A frequent supply of unique content is critical for building and sustaining an online following [2], and also represents one of the most challenging aspects of maintaining a single author blog. Group blogs are thus one of the best ways to minimize time investment and help maximize outreach efficiency: a group blog has the potential for more diversity (DSN includes female and LBGT voices) and reduces the burden on each individual. Group blogs can also invite posts from guest contributors, giving exposure to new scientists and helping to further broaden blog content. If you are interested in blogging, don't be afraid to ask an established blog about submitting a guest post—many sites welcome such contributions. A mixture of regular and guest contributors will naturally help to disseminate content more widely, since group blogs inherently leverage each person's own personal and professional networks. Regardless of the makeup of your blogging team, it is still vital to focus on publishing credible science using good communication skills.

## Rule 6: Focus on the Story

Producing something popular on the Internet is as much about passion and storytelling as it is about good content. The best content in the world, if delivered in a drab or ineffective manner, will not reach its desired audience or will fail to engage their attention. With passion, the right writing style, and deft use of digital media tools, you can make any type of science cool, and the importance of making things cool cannot be overstated.

Dickson [3] coined the term Information Deficit Model (also known as Science Deficit Model or just Deficit Model) for the notion that public mistrust of science results from a lack of understanding of scientific topics. The logical corollary of the IDM is that if we can simply overcome the knowledge deficit, public trust in science will improve. In our experience, there is no evidence that this actually happens. If we simply put the knowledge out there, most people will still lack the conceptual context to understand and acknowledge the significance of what they are being shown. We must not simply communicate the content of science, but also a passion for science. It is passion that is contagious and passion that drives scientists to push the frontiers of knowledge and understanding of the natural world. It will be passion that elevates the public to a greater appreciation of the transformative power of science; the knowledge deficit will take care of itself. Consider the success of the television series *Cosmos*, *The Undersea World of Jacques Cousteau*, and the BBC documentary works of Sir David Attenborough. In all three cases, there is incredible scientific content, to be sure, but the true success of these programs lies in their presenters—charismatic and authentic scientist/explorers who share their passion first and foremost, and the scientific content rides along with it.

One way that scientists can help convey passion is through storytelling. The notion of narrative structure is familiar to everyone, often unconsciously so, which provides great potential for scientific material that can be delivered in this fashion. Use of storytelling mechanisms employed by writers and journalists can help tremendously, as can explicit collaboration with artists, filmmakers, and other narrative experts. For writers, books such as "A Field Guide for Science Writers" [4] are invaluable resources for conducting effective outreach. In terms of visual media, there are many excellent outreach initiatives emerging on YouTube, such as "Minute Physics" (https://www.youtube.com/user/minutephysics), and "The Brain Scoop" produced by the Chicago Field Museum (https://www.youtube.com/user/thebrainscoop). RadioLab (http://www.radiolab.org/) is yet another enthralling example of scientific ideas conveyed in audio form, via podcast and public radio.

Another approach is to bridge the cultural gap between scientists and the public by explicitly creating commonalities with the reader. Scientists are not separate from the rest of society; we are also members of the public. We shop for groceries, visit the dry cleaners, take our kids to school, and vote in elections. We are influenced by society and engage in popular culture, and thus our communications and narratives can be deeply rooted in this idea. At DSN we aim to integrate pop culture with scientific content. This leverages virality and the vast exposure of pop culture phenomena, but also serves to show that the authors themselves are not aloof ivory tower-dwellers. For example, recent DSN posts have referenced Miley Cyrus (http://deepseanews.com/2014/02/mooring-family-photos/) and incorporated Internet memes (http://deepseanews.com/2013/09/lol-ocean-giants/).

## Rule 7: Leverage Multiple Tools to Disseminate Content and Build Up Your Network

We strongly encourage the use of multiple online tools, in order to reach different audiences and drive traffic to the main blog or website. Most readers tend to use one or two new media tools, with platform use depending on users' personal preferences and established online social networks. Thus, it is important to cross-promote new content to Tumblr, Twitter, Facebook, Pinterest, and other relevant platforms. At DSN, we use a division of labor for these types of cross-promotion, relying on authors that specialize in each different tool. We also try to automate as much of this process as possible, which is imperative for time management and efficiency in outreach activities. A suite of Wordpress plugins and standalone websites can be leveraged for automatically pushing content to different social media accounts (some examples include http://dlvr.it and http://twitterfeed.com), although such automation is best supplemented and balanced with human-led interaction with online audiences.

Actively engaging with audiences, not just broadcasting information, is also an important part of using these different tools. Kietzmann et al. [5] cite examples from the corporate world, where ineffective social media engagement (ignorance or misguided policies), can result in both missed opportunities and failure to mitigate potential bad press. For scientific outreach projects, monitoring accounts and participating in subsequent discussions (e.g., responding to Facebook/Blog comments, answering questions on Twitter) is a key component for sustaining participation from followers and encouraging growth of your online community. Social media engagement therefore helps audiences pursue their own interests, helps scientists address controversial topics like climate change, and simultaneously helps break down perceived barriers between scientists and society (see Rule 6).

Finally, we note that open licensing of online content can play a pivotal role in its dissemination. All content at Deep-Sea News is available under a Creative Commons (CC) license (http://creativecommons.org/), allowing users to freely share and redistribute the material with attribution to the original source. CC-licensed materials can be a long-term boon to outreach efforts, especially if content is widely shared and linked on sites such as Wikipedia. For example, the photo-sharing website Flickr, http://www.flickr.com, allows users to post and search for CC-licensed images. Data and figures can also be posted to repositories such as Figshare (http://figshare.com) where they are assigned a unique digital object identifier (DOI), making them both shareable and citable. Open content thus allows your outreach project to be built on and complemented by other people, by providing unique materials for audiences to engage with and share.

## Rule 8: Collect and Assess Data

Currently, there is much anecdotal and observational evidence regarding what scientists gain through the use of new media tools (in terms of professional benefits and outreach impacts), but not much in the way of systematic data and results. At Deep-Sea News, we attempt to collect as much data as feasible in order to gauge traffic spikes, content-related trends, and long-term growth in readership. Website metrics such as Google Analytics (http://www.google.com/analytics/) and Stat Counter (http://statcounter.com) are used to track site traffic, including the point of entry (page and/or search engine term), country of origin of visitors, and unique versus repeat visitors. Blog posts are specifically tailored with Search Engine Optimization plugins installed via Wordpress—each post has a category, keyword tags, and a descriptive blog title post that helps to drive search engine traffic. In addition, we keep track of social media metrics: number of tweets/retweets, Facebook "likes," and other shares of a given post.

Other strategies and metrics we use at DSN include blog comments and reader surveys. Reader comments are also an important measure of impact. Some blog posts inspire further conversation in the comments section that can extend to other social media platforms such as Twitter or Facebook. However, oftentimes the quantity of blog comments correlates to controversy more than popularity, and can be counterproductive to our science education mission. Therefore, to keep the comment section restricted to productive discussion, DSN has implemented clear moderation policies (http://deepseanews.com/about-2/commenting-policy/).

Despite this suite of metrics, as blog administrators we are often uncertain about how to best interpret and maximize the use of analytic data. For example, posts published at DSN for 2013 ($n$ = 299), garnered a total of 1,666,119 page views. Of these views, 82.6% were received on the top 20 posts; the lowest ranking 200 posts accounted for just 5% of total 2013 views. If anything, these trends illustrate the asymmetry of online reach and impact. Many posts receive a moderate amount of interest and reader comments, but occasional posts go viral and attract mainstream media attention, which has a dramatic positive effect on blog traffic. It is difficult to predetermine what posts will soar in popularity, but the longer you participate in effective online outreach, the more it is likely to happen. In this sense, reaching a truly large audience is a long game.

## Rule 9: Iteratively Assess What Works and What Doesn't

The above-mentioned analytics can be used to assess your online reach and track progress towards your goals. However, the online environment is still an evolving and untested sphere, and you will undoubtedly have to adopt a trial-and-error approach to find what works best. Reflecting on the 3,688 posts currently hosted at Deep-Sea News, we can provide some insight into the kinds of posts that work versus those that do not. Post length is an important aspect to consider, since different lengths can target different audiences and outreach goals. In general, posts that are well received on our site are usually 400–800 words, with liberal use of images and videos. Lists are particularly low-hanging fruit and often go viral (e.g., "Top Ten. . ." or "Best of.." posts). Occasional long posts (>800 words) can cater towards a more engaged, but smaller audience. Since these posts require more of the reader's time, they tend to be popular amongst people with an existing interest in science, or students and teachers seeking educational resources. Longer posts can also be quite effective in addressing public misconceptions, which often take considerable time to untangle.

Tone is a critical part of outreach identity. Successful writers often have a distinctive voice, which creates an interesting and engaging narrative. However, many writers must find their voice, rather than knowing it from the onset. The web is a fantastic place to try new writing styles, especially with near instant feedback in the form of Facebook likes and shares, tweets, or recognition from fellow bloggers. At DSN, we have found that humor is key. Humor, when used effectively, generates a relaxing and welcoming environment for the reader. Linking to pop culture or Internet memes also connects posts to a larger social context and can help people relate their lives to the science being discussed.

In the course of DSN's history, we have also realized that there are real barriers to good science outreach. First, crafting a good post is time-consuming, and can take many hours. By spreading the work around, group blogs help to alleviate some of the time pressure (Rule 5). Certain topics are also inherently more difficult to tackle, such as Fieldwork or Expedition blogging. In order to do this type of writing well, a strong "hook" and/or human interest is needed to draw the reader in—otherwise these types of posts are akin to sitting through a slideshow of someone else's travel photos. Controversial topics can also be difficult to address, as they often draw unwanted attention, criticism, and negativity to the blog—examples at DSN include fisheries, climate change, and the Sea Shepherds organization.

## Rule 10: Create Prestige for Public Scholarship

What do we truly gain from online outreach activities? As scientific professionals working in the research, academia, and nonprofit sectors, we are not evaluated on our outreach. However, we argue that there are a number of personal and professional benefits to be gained, as discussed in previous publications [1,6]. The most important overarching benefit is visibility—to one's colleagues, to the media, and to the public. By being accessible, researchers participating in online conversations have the opportunity to have a much more influential voice for their science. In these days of dwindling governmental investment and increased public distrust of science, scientists need to speak out on the value of their profession and training.

At DSN, we have derived professional benefits and personal satisfaction from our work, including published papers [1], new collaborations (e.g., http://deepseanews.com/2012/08/sharks-and-lasers-not-just-for-entertainment/), and substantial media coverage of our work (e.g., http://articles.latimes.com/2014/jan/12/local/la-me-west-coast-radiation-20140113). However, the most rewarding aspect is to have become an authoritative source on ocean science for the media and the public. Some of our writing, such as that on the Fukushima nuclear disaster and the Deepwater Horizon oil spill, has been widely quoted in the press, and our posts aimed at students interested in marine science are used by teachers and advisors around the world. Because we have witnessed such direct and beneficial gains as a result of our online outreach activities, we feel strongly that such activities should be given more weight when determining scientific productivity, e.g., during hiring/promotion decisions. The impact of online activities is increasingly recognized [7–8], and they should be formally encouraged.

## Conclusions

In the end, it's important to have reasonable expectations for your online activities. Don't be afraid to start small. Remember that not every single one of your posts will go viral—in fact, it will be very rare that they do. Online outreach is generally a long game. Content production and consistency are key factors that will impact how audiences view and access your blog.

Finally, quality and engagement are both important for becoming a trusted go-to source in the online world, and for extending your impact beyond the Internet.

## References


1. Bik HM, Goldstein MC (2013) An Introduction to Social Media for Scientists. PLoS Biol 11(4): e1001535. doi: 10.1371/journal.pbio.1001535 PMID: 23630451
2. Bailyn E (2012) Outsmarting Social Media: Profiting in the Age of Friendship Marketing. Indiana: Que Publishing. 192 p.
3. Dickson D (27 June 2005) The case for a 'deficit model' of science communication. SciDevNet. Available: http://www.scidev.net/global/communication/editorials/the-case-for-a-deficit-model-of-science-communic.html. Accessed 24 Mar 2014.
4. Blum D, Knudson M, Henig RM (2006) A field guide for science writers. Oxford: Oxford University Press. 336 p.
5. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS (2011) Social media? Get serious! Understanding the functional building blocks of social media. Bus Horiz 54(3): 241–251.
6. Darling E, Shiffman D, Côté I, Drew J (2013) The role of Twitter in the life cycle of a scientific publication. Ideas in Ecology and Evolution 6: 2–43.
7. Eysenbach G (2011) Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. J Med Internet Res 13(4): e123. doi: 10.2196/jmir.2012 PMID: 22173204
8. Priem J, Piwowar HA, Hemminger BM (2012) Altmetrics in the wild: Using social media to explore scholarly impact. Preprint. Available: arXiv:1203.4745. Accessed 24 Mar 2014.

Editorial

# Ten Simple Rules for Writing a PLOS Ten Simple Rules Article

**Harriet Dashnow[1,2,3]*, Andrew Lonsdale[4], Philip E. Bourne[5]**

1 Bioinformatics, Murdoch Childrens Research Institute, Parkville, Victoria, Australia, 2 Life Science Computation Centre, Victorian Life Sciences Computation Initiative, Carlton, Victoria, Australia, 3 The University of Melbourne, Parkville, Victoria, Australia, 4 The Australian Research Council Centre of Excellence in Plant Cell Walls, School of Botany, University of Melbourne, Parkville, Victoria, Australia, 5 Office of the Director, National Institutes of Health, Bethesda, Maryland United States of America

*[PB: When I read the title of this article I laughed out loud—how many times has that happened to you when reading professional articles? Laughter is good whatever the context. When I started the series in 2005, I had no idea it would be so successful. This article, which I had no part in writing, only adding commentary shown in italics, is in my mind a celebration of that success. My commentary is simply to provide a historical perspective to explain some aspects of why the collection is the way it is and, of course, to make a few personal observations which, after all, is what the collection is meant for. Thanks to HD and AL for making this happen and for including me as an author (see Rule 4) and to all those that have contributed over the past nine years.]*

## Introduction

Would Newton retweet your post on Twitter? How would Einstein view open source software? How would Darwin have handled a Wikipedia edit war?

The way we do science is changing almost as fast as the volume of our data. Advice is needed; however, advice on leading a successful scientific life is usually confined to outdated memoirs, unrecorded weekly lab meetings, neglected blogs, or casual conversations at a conference.

When we are faced with the challenges of how to be the best scientist we can, our instinctive reaction is to follow our usual pattern of inquiry—search the literature.

This search left us wanting, until we discovered the PLOS Ten Simple Rules collection. We have found them to be a series of concise articles that capture the professional zeitgeist of being a scientist in an approachable manner.

Many topics cover the professional (or "soft") skills that are necessary for a modern scientific career, but are not part of a formal scientific education. *[PB: Sad but true—teaching such skills should be a no-brainer.]* These articles represent an invaluable chance to pass on advice and knowledge in a way that can be widely distributed, formally recognised, and—as an added benefit—cited.

If (like us) you have read some articles in the Ten Simple Rules collection and appreciated their value, you may feel the urge to write one of your own. The collection provides a succinct and engaging format for advice on these skills. However, coming up with an article on soft skills need not be hard. Perhaps you have some insight, experience, or wisdom to impart. How would you do that?

Is there practical advice for contributing to the Ten Simple Rules collection already available? What can we learn from the existing articles in the collection? If only there was an article with ten simple rules for writing a PLOS Ten Simple Rules article. If only that article could be peppered with insightful comments from the founder of the collection: Philip E. Bourne.

This is that article.

## Rule 1: Have Ten Rules

Perhaps the most obvious prerequisite for writing an article entitled "Ten Simple Rules…" is to actually have ten rules (Figure 1). There can be a temptation to include unimportant points or to split one topic over multiple rules to get to that magic number ten. If you can't think of ten rules, maybe your topic is too specific for this format.

Another common problem is having too much to talk about. This one is a little easier to deal with. Don't be tempted to emulate Spinal Tap by "turning it up to eleven." Simply rank your ideas by how important or how thought provoking they are and then just write about the top ten. Alternatively, if you have twenty rules and there is a clear split, you may have two articles on your hands. There may be scope to combine a few related ideas within one rule, but don't get carried away. Your Ten Simple Rules should be just that, simple.

*[PB: Surprisingly, in the Ten Simple Rules articles I have written, I have found the imposition of ten never to be an issue—somehow it has seemed just enough, but not too much. Undoubtedly important points have been omitted but what is there seems to hang together. When authors suggest a topic for the Ten Rules I say send the Rules first no text. If they stand alone and say something new then I encourage them to flesh them out.]*

## Rule 2: Choose Your Topic Wisely

The articles in the Ten Simple Rules collection share an almost intangible common component: everything you always wanted to know about science (but perhaps were afraid to ask). These are articles about how to get by in the world of scientific research. Some give specific guidance in the field of bioinformatics and computational biology (naturally, as the collection originated in *PLOS Computational Biology*), but most offer broad advice that reaches far beyond this demographic.

Consider the topics covered so far: advice to graduate students, getting the right postdoctoral position, choosing be-

**Citation:** Dashnow H, Lonsdale A, Bourne PE (2014) Ten Simple Rules for Writing a PLOS Ten Simple Rules Article. PLoS Comput Biol 10(10): e1003858. doi:10.1371/journal.pcbi.1003858

**Editor:** Fran Lewitter, Whitehead Institute, United States of America

**Published** October 23, 2014



**Funding:** The authors received no specific funding for writing this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* Email: harriet.dashnow@unimelb.edu.au

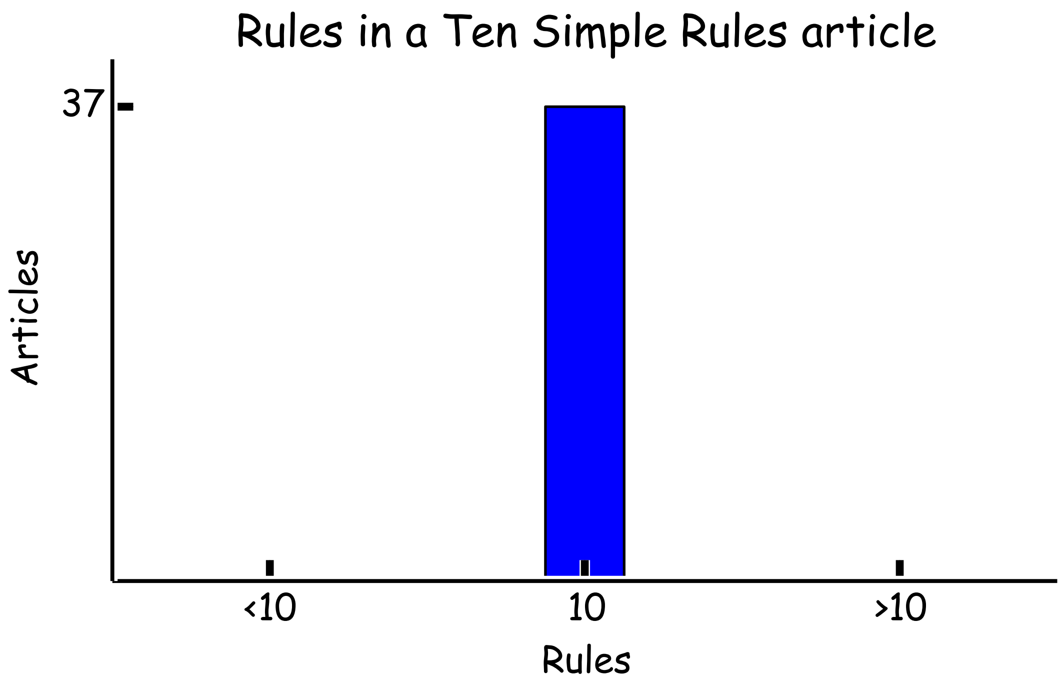



**Figure 1. Have ten rules.**
doi:10.1371/journal.pcbi.1003858.g001

tween academia and industry, and how to start a company. The articles even extend to interpersonal skills like networking, collaboration, and communicating with your supervisor. There is also a strong academic focus: the core skills of doing research, writing, publishing, and teaching.

Now, what can you add to the mix? What do you wish that you'd known a year ago? What would you tell a new student? What do you wish your supervisor would realise? Computational biology is such a fresh, fast-moving field. There is always room for advice to emerge and for those new to the field to share their experiences, drawing from other disciplines and bringing together emerging ideas and techniques. The Ten Simple Rules collection is a reservoir for accessible wisdom (in both the open source and intellectually approachable senses of the word). What do you have to contribute?

If you are stumped, here are some suggestions: Ten Simple Rules for retiring when you know you should. Ten Simple Rules for winning a Nobel Prize. Ten Simple Rules for managing a scientific rivalry.

*[PB: Some would say I should write the first; none would say I should write the second; the third I have enough of already, but having an entry in the series from a Nobel laureate is a great idea—I am working on recruiting that someone right now—and so the series goes.]*

## Rule 3: Include an Introduction

Never underestimate the importance of a great introduction. Your introduction defines the scope of your article. It sets up a promise to your reader that you will cover this, that, but not some other thing. It provides the opportunity to put your topic in perspective and fill any gaps in your reader's knowledge.

Perhaps the most vital function of the introduction is to convince your reader to keep reading. You need to make your point clearly and simply the first time. You are also going to need to catch their attention in the first paragraph or so. You must convince them that you can both entertain and inform.

Like your Ten Simple Rules, your introduction should be short and to the point—there is no point having Ten Simple Rules with a thesis-length introduction.

*[PB: Agreed, particularly concerning the length of the introduction.]*

## Rule 4: Be Philip E. Bourne

Your best chance of having a PLOS Ten Simple Rules article published is by being Philip E. Bourne. At the time of writing, 48.6% of PLOS Ten Simple Rules articles list *PLOS Computational Biology* Founding Editor-in-Chief, Philip E. Bourne, as the first, last, or sole author (Figure 2). In fact, statistically speaking (Box 1), his frequent occurrences as an author of Ten Simple Rules is significant (p-value = 6.63758e-52).

More recent articles are increasingly written by other experts or groups of individuals, but Philip is often found as part of the author list. Before you start filling out a change of name form though, consider asking Philip to be a coauthor.

*[PB: And I thought they asked me because I know something.]*

## Rule 5: Collaborate

You have a great idea for a Ten Simple Rules article, now you need to make it a reality. We have discussed getting Philip Bourne involved (see Rule 4), but we haven't yet discussed why having a coauthor or two could help, even if they don't have an in at PLOS.

When you write a Ten Simple Rules article, you are speaking with an aura of knowledge and authority. One way to achieve this is to be a giant in your field. But if only the "experts" write these articles, we are missing out on the intimate, hard-earned knowledge of those in the trenches. If your name doesn't strike awe in your readers, perhaps consider combining forces with others. If you can't find experts, contact your peers. The emerging opinion of a group of more junior authors may be almost as respect-

**Box 1. Philip E. Bourne Is Significantly Over-Represented as an Author in the PLOS Ten Simple Rules Collection**

Showing the methods in the main text for the statistical significance of being Philip E. Bourne went against the flow of our article, and might have scared readers off. Therefore we've hidden them here to show that even in a Ten Simple Rules article, you need to back up your statistical claims.

We consider Bourne as an author in the Ten Simple Rules collection through over-representation analysis, where we look for over-representation of authorship in our group of interest, the Ten Simple Rules collection, compared to the overall PLOS collection.

Let: $k = $ Number of Ten Simple Rules articles where Bourne is an author, $n = $ Total number of Ten Simple Rules articles, $K = $ Number of PLOS articles where Bourne is an author, and $N = $ Total number of PLOS articles.

Then $\Pr(X = k) \sim$ hypergeometric.

$k = 18$, $n = 37$, $K = 55$, $N = 119435$

```
Pr(x>=k) = phyper(k-1, K, N-K, n, lower.tail=FALSE) # R code
```

$= 6.63758\text{e-}52$

Bourne is significantly over-represented as an author in the Ten Simple Rules collection (compared to what we would expect by chance based on his publication rate in the entire PLOS archive).

We speculate that as the Founding Editor-in-Chief of *PLOS Computational Biology* and founder of the series, Bourne is positively disposed towards publishing articles in the collection, and this disposition accounts for the high number of Ten Simple Rules articles he has authored.

Further experiments are required to validate this theory. Unfortunately, ethics approval to experiment on Bourne was not forthcoming. Bourne himself registered strong objections to our proposed ''knock out'' tests.

*[PB: He has been knocked around enough already over his career—wait that could be a topic for a new Ten Simple Rules article.]*

able, particularly if the topic at hand is relevant.

Crowdsourcing your peers is also a great way to find the right ten rules. Gather rules from a number of people and look at the intersection. This may give you a sense of the community consensus rather than an individual opinion.

*[PB: Agree that everyone has something to offer the collection. With respect to having "an in at PLOS," I should say that all contributions for which I am not an author have me as an editor and are sent for review. Another of the senior editors handles those that have me as an author. Either way, they often require significant editing and, on occasion, are rejected.]*

## Rule 6: Research

When writing any academic paper, reading the literature is a given. So, read the Ten Simple Rules collection. There is no excuse: there are only ten rules per paper!

After you have read this article, and some of the other articles in the collection, you will have a good idea of the required style and tone.

To show that we can take our own advice, we first conducted a thorough review of the Ten Simple Rules, culminating in the first draft of this article.

To gain a more historical perspective we also searched the literature for titles containing "ten simple rules."

The first entry in the PLOS Ten Simple Rules collection was published in 2005. The earliest entry matching our search was published in 1988 ("Ten simple rules for improving advertising in health care institutions") and though the colloquial origin of the phrasing is beyond the scope of this paper, it is the first example we can find of a journal article that fits the format later used by the PLOS collection.

Articles within the PLOS collection have inspired articles in other fields, including blog posts, contributions to preprint servers, and journals such as *NeuroImage* and the *International Journal for Parasitology: Parasites and Wildlife*.

The *NeuroImage* branch of ten simple rules articles, with a length, focus, and complexity that may call into question the use of the term "simple," demonstrates how the ten simple rules format can be adapted to suit any need and discipline. The authors of this review resisted the urge to produce a phylogenetic tree. Just.

*[PB: I certainly make no claim to be the originator of the phrase "Ten Simple Rules." It just seemed to fit the length and form of what was required to get the point across in that very first article, and clearly others thought the same as additions to the series came rolling in. The rumour that the appeal of "simple" is that each entry in the series is the right length for a bathroom or lavatory break should be disregarded.]*

## Rule 7: Write Well

The topic of writing well deserves its own ten simple rules. There are plenty of great online resources about how to write well. We won't try to enumerate the rules here, but instead will focus on one of the most important: know your audience.

So, who reads *PLOS Computational Biology*? Who reads the Ten Simple Rules articles specifically?

The audience is highly educated and has a great deal of specialist expertise. They may be skimming this article over a coffee in the middle of checking their email in the morning. Unlike a scientific article, the reader of a Ten Simple Rules article is unlikely to read it twice for comprehension, or to try to decipher any complex language.

There is no single writing style that you can successfully apply to every domain. To select your style, you must first know who you are writing for. Who is your audience? What are they interested in? What do they already know about this topic? How much time are they likely to spend reading this article?

If you can answer these questions, you are a long way towards getting the tone and content right.

*[PB: Wise words that in one way or another are stated in many of the articles in the series and so it must be true. In terms of style we have been flexible—if the style is different, but works for the content, so be it. This article is, of course, a case in point.]*

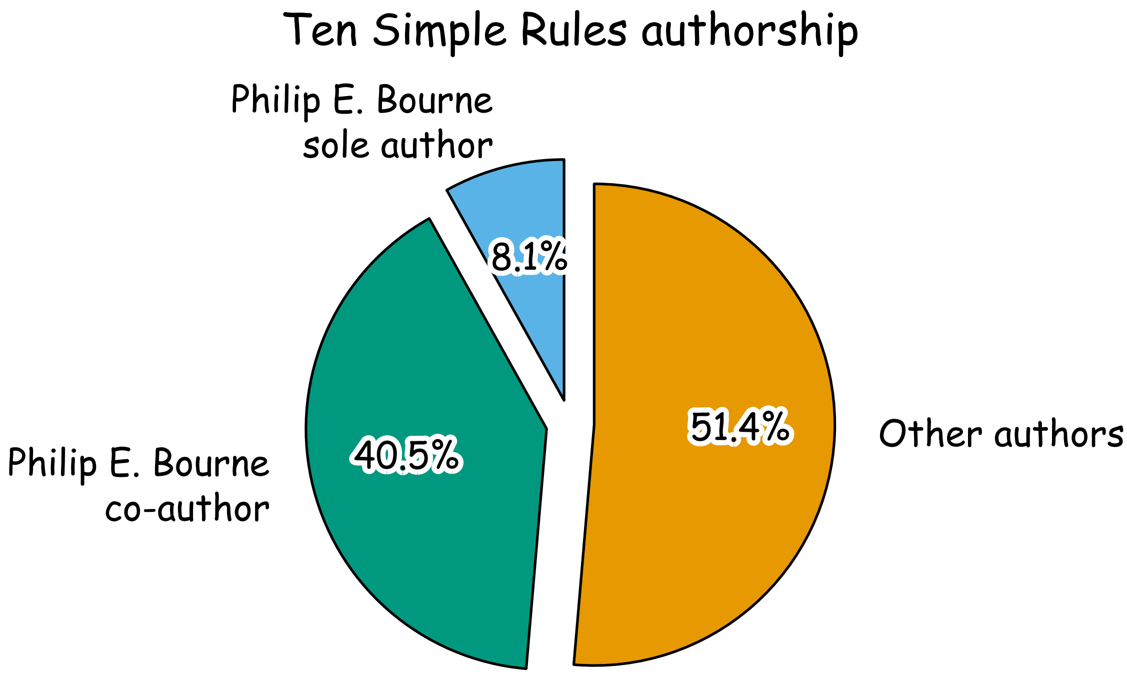



**Figure 2. Your best chance of having a PLOS Ten Simple Rules article published is by being Philip E. Bourne.**
doi:10.1371/journal.pcbi.1003858.g002

## Rule 8: Reference

The ten rules format calls for a casual writing style, but that shouldn't prevent you from referencing. The need for citations will vary with the topic, however this is still a scientific journal, and this paper will form part of the scientific literature, indexed and found. There is plenty of scope for presenting your opinion, but do back up your facts. Wherever possible cite relevant online resources, including other Ten Simple Rules articles.

The caveat to this rule is that there can be such a thing as too many references. A great Ten Simple Rules article is easy to read and accessible to its audience. There is little point in having ten simple rules then overloading them with hundreds of references. Include only those that you have read and learnt from.

Resolve this conflict by considering how your article will be used. How many people will read it? Although references are important, Ten Simple Rules articles are viewed and downloaded more often than they are cited (Figure 3). In fact, by recent count, six of the top ten most viewed articles in *PLOS Computational Biology* were from the Ten Simple Rules collection. Clearly, the impact of your article will not be measured by citations.

By considering your audience (see Rule 7), you will get a sense for the number of citations needed to effectively get your message across. Balance accessibility, brevity and authority, as these qualities that will determine how many people have a chance to follow your ten simple rules.

*[PB: I love this graph (Figure 3). What does it say about impact? After all is this not what we are trying to measure? PLOS's efforts with article level metrics (ALMs) speak to the need to be more quantitative in how we measure the impact of a piece of scholarship. But since what we do most of the time is be qualitative—judging a piece of work by the impact factor of the whole journal—let me also be quantitative. Judging by the number of times someone I do not know comes up to me and says, "I know your work" and I respond, "really which research are you referring to?" Upon which they say, "I don't know your research, I am talking about the Ten Simple Rules," I would say the impact is high—at least relative to my research.]*

## Rule 9: Edit

Once you have the words on the page, your next step is to edit them. Present your rules in a logical order. Order and reorder them. Find the flow both between and within your rules.

Once you have taken care of the big picture, spend some time thinking about whether your paragraphs are cohesive and well structured. This is not an essay, so the rules are a little more fluid, but you could still benefit from topic sentences and phrases that flow between paragraphs. If, like many academics, you find your sentences blow out beyond 30 words, you may like to use a tool such as Draft to keep them in check. These tools are great for picking up tortured sentence structures like passive voice and split infinitives.

One of the best ways to make your article more readable is to have others read it. Workshop and crowdsource your article. Your peers are your secret weapons. Many of them are your target audience, so you can do a test run and perhaps even get honest feedback.

Some literary theorists argue that as soon as you publish a piece of writing, it becomes the domain of the reader. The intent of the author pales in comparison to the interpretation of the audience. In this context, the reader is always right. So you may as well get the major criticisms and revisions out of the way early.

*[PB: As I say somewhere in the collection, what you commit to the literature will be around long after you are gone—it is a large part of your professional legacy—get it right. Reviewers' comments often relate to the order and organization of the rules, so this rule is good advice for getting your contribution published and out there.]*

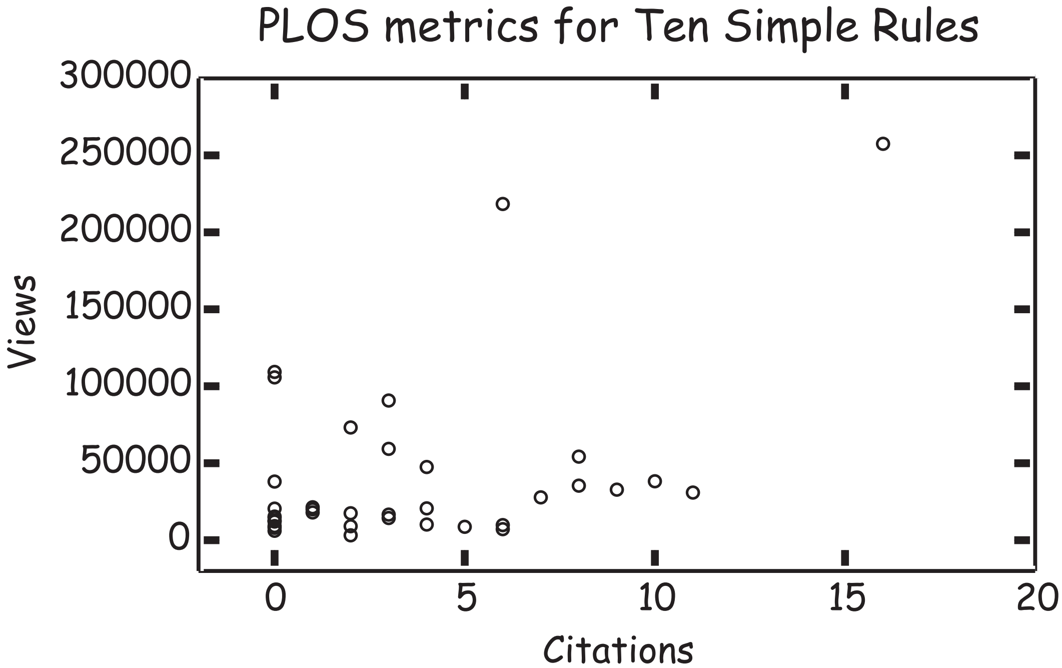



**Figure 3. Ten Simple Rules articles are viewed more often than they are cited.**
doi:10.1371/journal.pcbi.1003858.g003

## Rule 10: Have a Voice

Inject your personality. The coauthors of this article have included several jokes that did not make it past the editing stage. There was a joke about "soft skills informatics" and "professional development-omics" that was rightly cut, but that is part of the process. There are few scientific articles where you can borrow your plot style from the XKCD comic series (http://xkcd.com/), and this article is one of them.

That said, avoid in-jokes. Although perhaps incredibly funny to you, they don't tend to translate well to a wider audience. Your jokes should make most of your readers feel like they are part of the inner circle.

Does your voice match your article? In this article we have strived to balance humour with useful advice.

Our voice has mixed facts with seemingly irrelevant plots and calculations, and has attempted to convey a clear affection for the Ten Simple Rules collection, and shared useful tips to inspire you to write your own Ten Simple Rules.

*[PB: Personally I think it a shame that scientific discourse is taken more seriously the more impersonal it is. Why should humour and individuality impact the perceived value of science? With the Ten Simple Rules, you have the opportunity to buck that system as this article so rightly illustrates.]*

## Conclusion

You should definitely have a conclusion. Many readers will read the introduction, Rule 1, and then (if you're lucky) skim until they reach the conclusion. This is your opportunity to present a take-home message to your readers. Something to make sense of the rambling mess that is your article, despite extensive editing. Something to make it seem focused and insightful. It's also a great opportunity to re-inject your personality into your writing.

So here is our conclusion:

Congratulations on making it this far!

Hopefully we've convinced you that the Ten Simple Rules collection has a vital role to play in modern science. We'd like you to think about contributing to the discussion.

When you do, take care to choose your topic wisely. Success in this genre is all about knowing your audience. Decipher what they are interested in and how they interact with these articles. A great way of keeping your reader involved is to use your voice, play up your persona, and, of course, don't forget to have ten rules.

The rules listed in this article are a guide to forming your own ten simple rules. These rules are simple, but not trivial. Use this article as a guide, and get started. Today. Just take it one rule at a time.

@Newton, can I get a RT?

*[PB: I would like to think the articles I have written come from the heart, a genuine desire to short-circuit all the mistakes I have made in a long career. I think that desire to share your experiences so others can learn should be what guides you in writing a Ten Simple Rules article. I look forward to continuing to read your efforts.]*

## Supporting Information

**Text S1** Data and code used to produce figures.
(DOCX)


## Acknowledgments

It's unlikely the advice you are dispensing has arisen in a vacuum. Give credit. This review article itself was inspired by a journal club meeting devoted to the Ten Simple Rules collection, and the authors would like to thank all the members of the Parkville Bioinformatics Journal Club (http://parkville-bioinformatics-journal-club.blogspot.com.au/) for their feedback, discussion, and encouragement. Special thanks to Danielle Ingle, Lesley Raven, Michael Walker, Scott Ritchie, Marek Cmero, and Jocelyn Penington for their invaluable comments. Thank you Alicia Oshlack for your advice, support, and comments on the manuscript. This article was written collaboratively during the fortnightly Parkville Bioinformatics Journal Club writing gym sessions using the Draft (https://draftin.com/) writing tool.

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Better Figures

**Nicolas P. Rougier[1,2,3]*, Michael Droettboom[4], Philip E. Bourne[5]**

1 INRIA Bordeaux Sud-Ouest, Talence, France, 2 LaBRI, UMR 5800 CNRS, Talence, France, 3 Institute of Neurodegenerative Diseases, UMR 5293 CNRS, Bordeaux, France, 4 Space Telescope Science Institute, Baltimore, Maryland, United States of America, 5 Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America

Scientific visualization is classically defined as the process of graphically displaying scientific data. However, this process is far from direct or automatic. There are so many different ways to represent the same data: scatter plots, linear plots, bar plots, and pie charts, to name just a few. Furthermore, the same data, using the same type of plot, may be perceived very differently depending on who is looking at the figure. A more accurate definition for scientific visualization would be a graphical interface between people and data. In this short article, we do not pretend to explain everything about this interface; rather, see [1,2] for introductory work. Instead we aim to provide a basic set of rules to improve figure design and to explain some of the common pitfalls.

## Rule 1: Know Your Audience

Given the definition above, problems arise when how a visual is perceived differs significantly from the intent of the conveyer. Consequently, it is important to identify, as early as possible in the design process, the audience and the message the visual is to convey. The graphical design of the visual should be informed by this intent. If you are making a figure for yourself and your direct collaborators, you can possibly skip a number of steps in the design process, because each of you knows what the figure is about. However, if you intend to publish a figure in a scientific journal, you should make sure your figure is correct and conveys all the relevant information to a broader audience. Student audiences require special care since the goal for that situation is to explain a concept. In that case, you may have to add extra information to make sure the concept is fully understood. Finally, the general public may be the most difficult audience of all since you need to design a simple, possibly approximated, figure that reveals only the most salient part of your research (Figure 1). This has proven to be a difficult exercise [3].

## Rule 2: Identify Your Message

A figure is meant to express an idea or introduce some facts or a result that would be too long (or nearly impossible) to explain only with words, be it for an article or during a time-limited oral presentation. In this context, it is important to clearly identify the role of the figure, i.e., what is the underlying message and how can a figure best express this message? Once clearly identified, this message will be a strong guide for the design of the figure, as shown in Figure 2. Only after identifying the message will it be worth the time to develop your figure, just as you would take the time to craft your words and sentences when writing an article only after deciding on the main points of the text. If your figure is able to convey a striking message at first glance, chances are increased that your article will draw more attention from the community.

## Rule 3: Adapt the Figure to the Support Medium

A figure can be displayed on a variety of media, such as a poster, a computer monitor, a projection screen (as in an oral presentation), or a simple sheet of paper (as in a printed article). Each of these media represents different physical sizes for the figure, but more importantly, each of them also implies different ways of viewing and interacting with the figure. For example, during an oral presentation, a figure will be displayed for a limited time. Thus, the viewer must quickly understand what is displayed and what it represents while still listening to your explanation. In such a situation, the figure must be kept simple and the message must be visually salient in order to grab attention, as shown in Figure 3. It is also important to keep in mind that during oral presentations, figures will be video-projected and will be seen from a distance, and figure elements must consequently be made thicker (lines) or bigger (points, text), colors should have strong contrast, and vertical text should be avoided, etc. For a journal article, the situation is totally different, because the reader is able to view the figure as long as necessary. This means a lot of details can be added, along with complementary explanations in the caption. If we take into account the fact that more and more people now read articles on computer screens, they also have the possibility to zoom and drag the figure. Ideally, each type of support medium requires a different figure, and you should abandon the practice of extracting a figure from your article to be put, as is, in your oral presentation.

## Rule 4: Captions Are Not Optional

Whether describing an experimental setup, introducing a new model, or presenting new results, you cannot explain everything within the figure itself—a figure should be accompanied by a caption. The caption explains how to read the figure and provides additional precision for what cannot be graphically represented. This can be thought of as the explanation you would give during an oral presentation, or in front of a poster, but with the difference that you must think in advance about the questions people would ask. For example, if you have a bar plot, do not expect the

**Citation:** Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLoS Comput Biol 10(9): e1003833. doi:10.1371/journal.pcbi.1003833

**Editor:** Scott Markel, Accelrys, United States of America

**Published** September 11, 2014



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* Email: Nicolas.Rougier@inria.fr

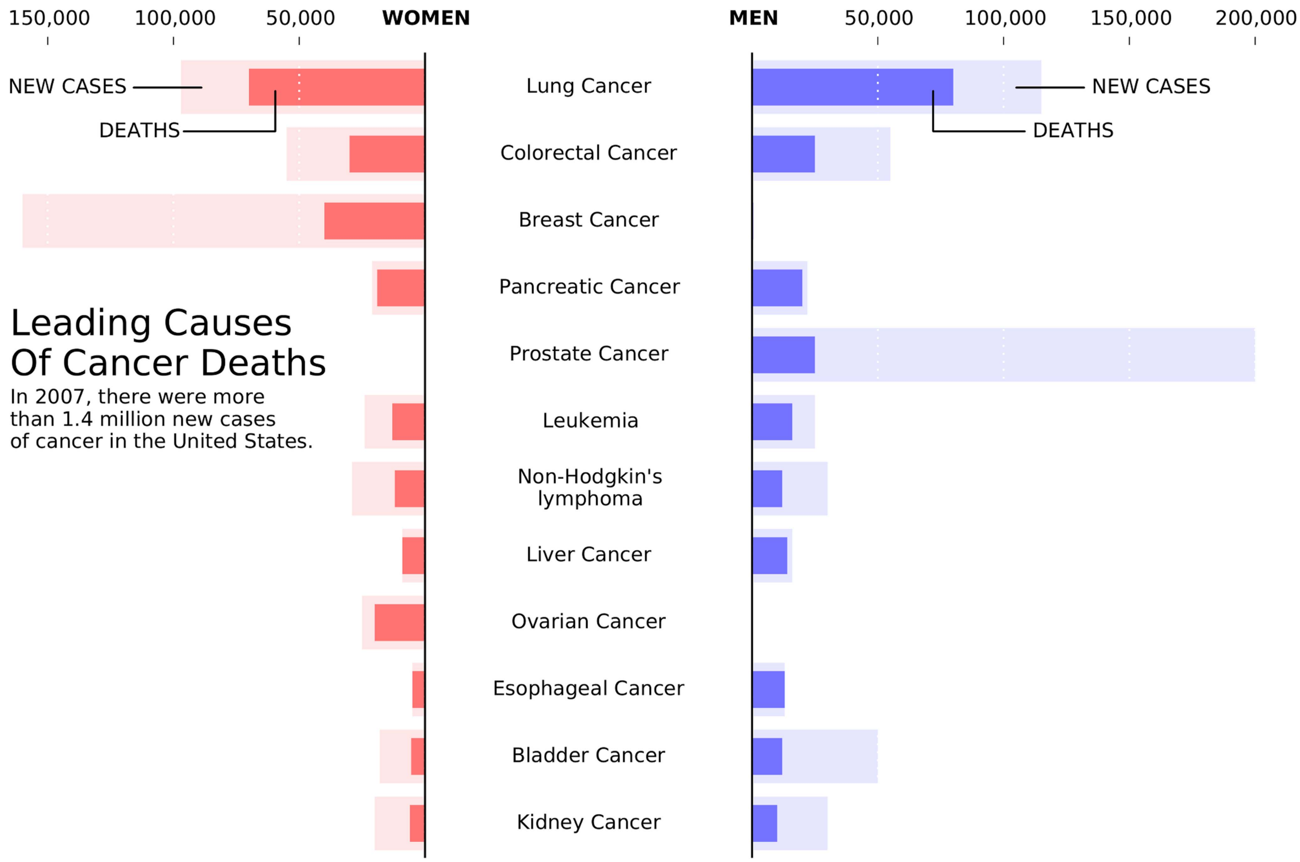



**Figure 1. Know your audience.** This is a remake of a figure that was originally published in the New York Times (NYT) in 2007. This new figure was made with matplotlib using approximated data. The data is made of four series (men deaths/cases, women deaths/cases) that could have been displayed using classical double column (deaths/cases) bar plots. However, the layout used here is better for the intended audience. It exploits the fact that the number of new cases is always greater than the corresponding number of deaths to mix the two values. It also takes advantage of the reading direction (English [left-to-right] for NYT) in order to ease comparison between men and women while the central labels give an immediate access to the main message of the figure (cancer). This is a self-contained figure that delivers a clear message on cancer deaths. However, it is not precise. The chosen layout makes it actually difficult to estimate the number of kidney cancer deaths because of its bottom position and the location of the labelled ticks at the top. While this is acceptable for a general-audience publication, it would not be acceptable in a scientific publication if actual numerical values were not given elsewhere in the article.
doi:10.1371/journal.pcbi.1003833.g001

reader to guess the value of the different bars by just looking and measuring relative heights on the figure. If the numeric values are important, they must be provided elsewhere in your article or be written very clearly on the figure. Similarly, if there is a point of interest in the figure (critical domain, specific point, etc.), make sure it is visually distinct but do not hesitate to point it out again in the caption.

## Rule 5: Do Not Trust the Defaults

Any plotting library or software comes with a set of default settings. When the end-user does not specify anything, these default settings are used to specify size, font, colors, styles, ticks, markers, etc. (Figure 4). Virtually any setting can be specified, and you can usually recognize the specific style of each software package (Matlab, Excel, Keynote, etc.) or library (LaTeX, matplotlib, gnuplot, etc.) thanks to the choice of these default settings. Since these settings are to be used for virtually any type of plot, they are not fine-tuned for a specific type of plot. In other words, they are good enough for any plot but they are best for none. All plots require at least some manual tuning of the different settings to better express the message, be it for making a precise plot more salient to a broad audience, or to choose the best colormap for the nature of the data. For example, see [4] for how to go from the default settings to a nicer visual in the case of the matplotlib library.

## Rule 6: Use Color Effectively

Color is an important dimension in human vision and is consequently equally important in the design of a scientific figure. However, as explained by Edward Tufte [1], color can be either your greatest ally or your worst enemy if not used properly. If you decide to use color, you should consider which colors to use and where to use them. For example, to highlight some element of a figure, you can use color for this element while keeping other elements gray or black. This provides an enhancing effect. However, if you have no such need, you need to ask yourself, "Is there any reason this plot is blue and not black?" If you don't know the answer, just keep it black. The same holds true for colormaps. Do not use the default colormap (e.g., jet or rainbow)

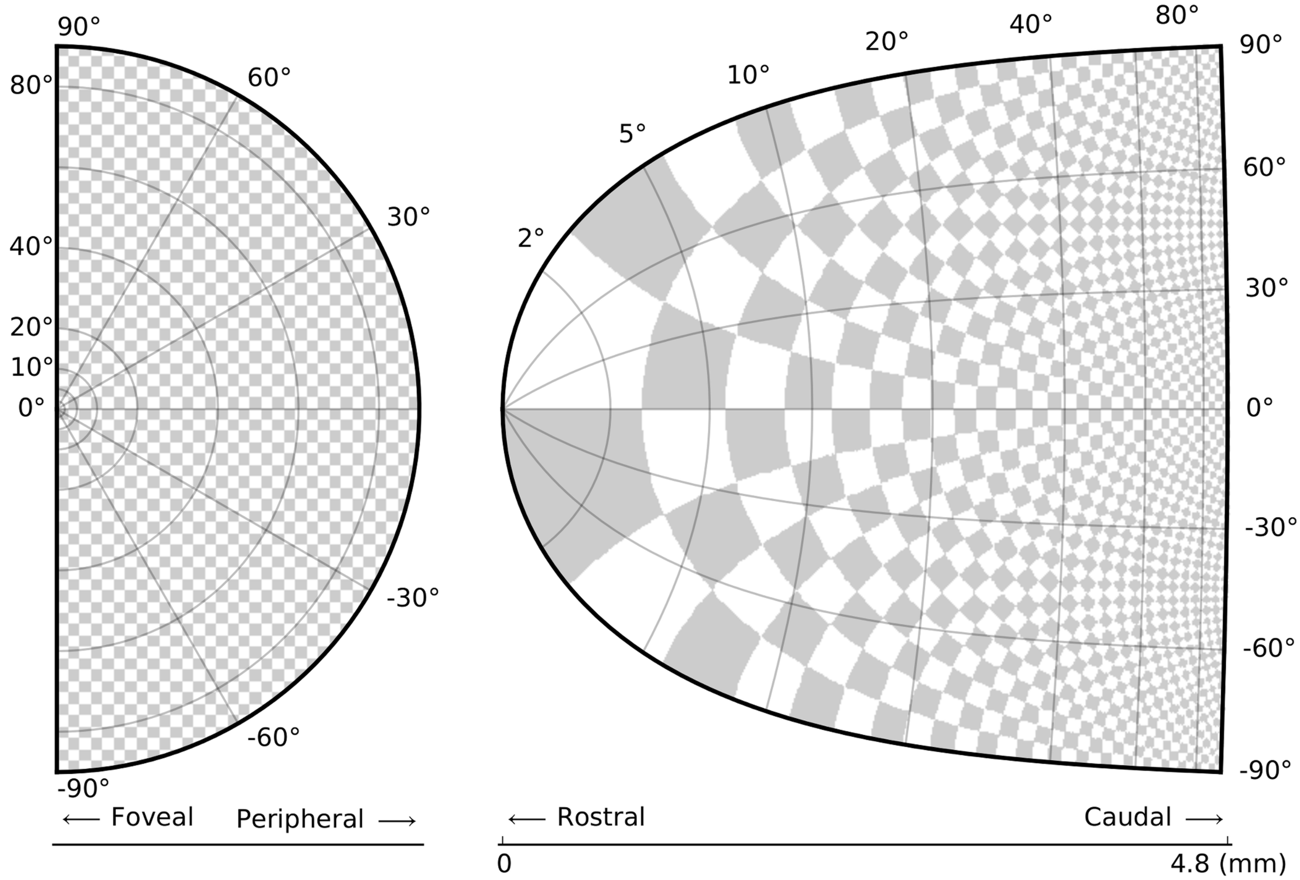



**Figure 2. Identify your message.** The superior colliculus (SC) is a brainstem structure at the crossroads of multiple functional pathways. Several neurophysiological studies suggest that the population of active neurons in the SC encodes the location of a visual target that induces saccadic eye movement. The projection from the retina surface (on the left) to the collicular surface (on the right) is based on a standard and quantitative model in which a logarithmic mapping function ensures the projection from retinal coordinates to collicular coordinates. This logarithmic mapping plays a major role in saccade decision. To better illustrate this role, an artificial checkerboard pattern has been used, even though such a pattern is not used during experiments. This checkerboard pattern clearly demonstrates the extreme magnification of the foveal region, which is the main message of the figure.
doi:10.1371/journal.pcbi.1003833.g002

unless there is an explicit reason to do so (see Figure 5 and [5]). Colormaps are traditionally classified into three main categories:

- *Sequential*: one variation of a unique color, used for quantitative data varying from low to high.
- *Diverging*: variation from one color to another, used to highlight deviation from a median value.
- *Qualitative*: rapid variation of colors, used mainly for discrete or categorical data.

Use the colormap that is the most relevant to your data. Lastly, avoid using too many similar colors since color blindness may make it difficult to discern some color differences (see [6] for detailed explanation).

## Rule 7: Do Not Mislead the Reader

What distinguishes a scientific figure from other graphical artwork is the presence of data that needs to be shown as objectively as possible. A scientific figure is, by definition, tied to the data (be it an experimental setup, a model, or some results) and if you loosen this tie, you may unintentionally project a different message than intended. However, representing results objectively is not always straightforward. For example, a number of implicit choices made by the library or software you're using that are meant to be accurate in most situations may also mislead the viewer under certain circumstances. If your software automatically rescales values, you might obtain an objective representation of the data (because title, labels, and ticks indicate clearly what is actually displayed) that is nonetheless visually misleading (see bar plot in Figure 6); you have inadvertently misled your readers into visually believing something that does not exist in your data. You can also make explicit choices that are wrong by design, such as using pie charts or 3-D charts to compare quantities. These two kinds of plots are known to induce an incorrect perception of quantities and it requires some expertise to use them properly. As a rule of thumb, make sure to always use the simplest type of plots that can convey your message and make sure to use labels, ticks, title, and the full range of values when relevant. Lastly, do not hesitate to ask colleagues about their interpretation of your figures.

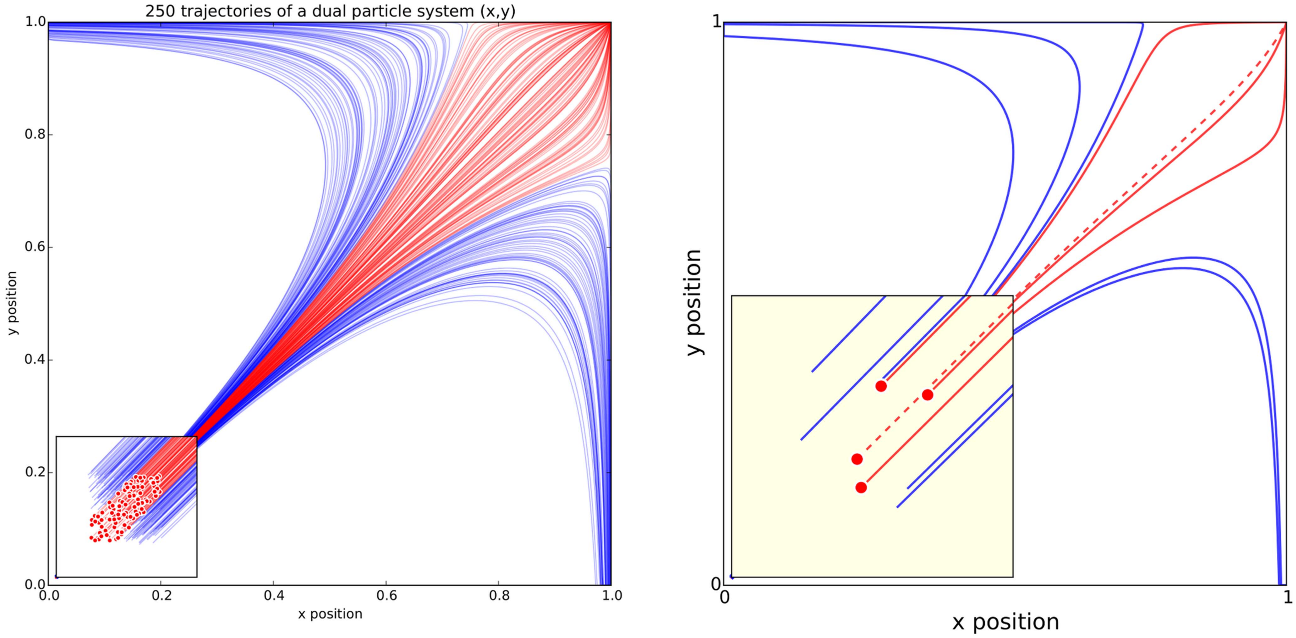



**Figure 3. Adapt the figure to the support medium.** These two figures represent the same simulation of the trajectories of a dual-particle system ($\frac{dx}{dt} = (^1\!/\!_4 + (x-y))(1-x),\ x \geq 0,\ \frac{dy}{dt} = (^1\!/\!_4 + (y-x))(1-y),\ y \geq 0$) where each particle interacts with the other. Depending on the initial conditions, the system may end up in three different states. The left figure has been prepared for a journal article where the reader is free to look at every detail. The red color has been used consistently to indicate both initial conditions (red dots in the zoomed panel) and trajectories (red lines). Line transparency has been increased in order to highlight regions where trajectories overlap (high color density). The right figure has been prepared for an oral presentation. Many details have been removed (reduced number of trajectories, no overlapping trajectories, reduced number of ticks, bigger axis and tick labels, no title, thicker lines) because the time-limited display of this figure would not allow for the audience to scrutinize every detail. Furthermore, since the figure will be described during the oral presentation, some parts have been modified to make them easier to reference (e.g., the yellow box, the red dashed line).
doi:10.1371/journal.pcbi.1003833.g003

## Rule 8: Avoid "Chartjunk"

Chartjunk refers to all the unnecessary or confusing visual elements found in a figure that do not improve the message (in the best case) or add confusion (in the worst case). For example, chartjunk may include the use of too many colors, too many labels, gratuitously colored backgrounds, useless grid lines, etc. (see left part of Figure 7). The term was first coined by Edward Tutfe in [1], in which he argues that any decorations that do not tell the viewer something new must be banned: "Regardless of the cause, it is all non-data-ink or redundant data-ink, and it is often chartjunk." Thus, in order to avoid chartjunk, try to save ink, or electrons in the computing era. Stephen Few reminds us in [7] that graphs should ideally "represent all the data that is needed to see and understand what's meaningful." However, an element that could be considered chartjunk in one

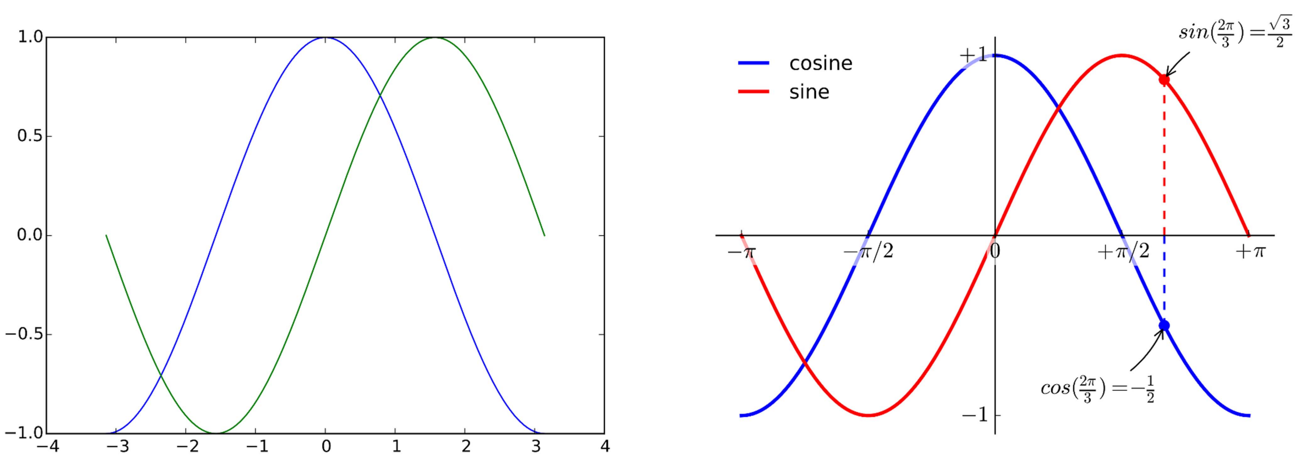



**Figure 4. Do not trust the defaults.** The left panel shows the sine and cosine functions as rendered by matplotlib using default settings. While this figure is clear enough, it can be visually improved by tweaking the various available settings, as shown on the right panel.
doi:10.1371/journal.pcbi.1003833.g004

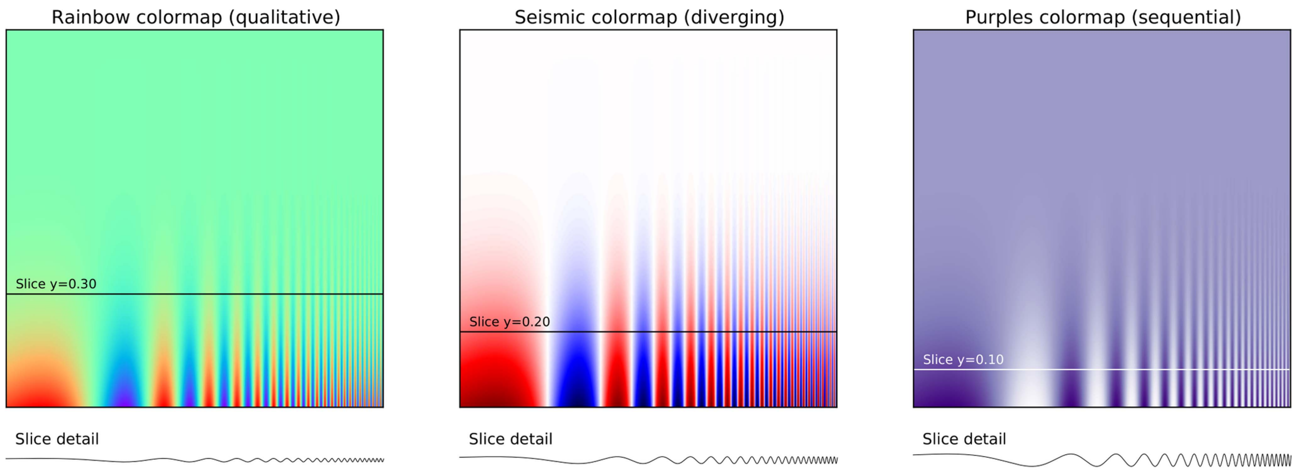



**Figure 5. Use color effectively.** This figure represents the same signal, whose frequency increases to the right and intensity increases towards the bottom, using three different colormaps. The rainbow colormap (qualitative) and the seismic colormap (diverging) are equally bad for such a signal because they tend to hide details in the high frequency domain (bottom-right part). Using a sequential colormap such as the purple one, it is easier to see details in the high frequency domain. Adapted from [5].
doi:10.1371/journal.pcbi.1003833.g005

figure can be justified in another. For example, the use of a background color in a regular plot is generally a bad idea because it does not bring useful information. However, in the right part of Figure 7, we use a gray background box to indicate the range [−1,+1] as described in the caption. If you're in doubt, do not hesitate to consult the excellent blog of Kaiser Fung [8], which explains quite clearly the concept of chartjunk through the study of many examples.

## Rule 9: Message Trumps Beauty

Figures have been used in scientific literature since antiquity. Over the years, a lot of progress has been made, and each scientific domain has developed its own set of best practices. It is important to know these standards, because they facilitate a more direct comparison between models, studies, or experiments. More importantly, they can help you to spot obvious errors in your results. However, most of the time, you may need to design a brand-new figure, because there is no standard way of describing your research. In such a case, browsing the scientific literature is a good starting point. If some article displays a stunning figure to introduce results similar to yours, you might want to try to adapt the figure for your own needs (note that we did not say copy; be careful with image copyright). If you turn to the web, you have to be very careful, because the frontiers between data visualization, info-

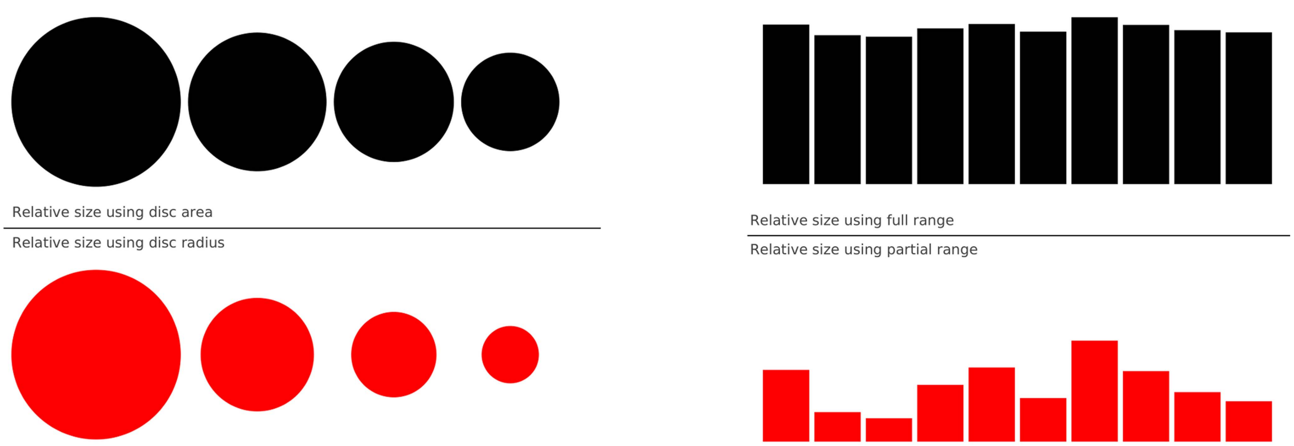



**Figure 6. Do not mislead the reader.** On the left part of the figure, we represented a series of four values: 30, 20, 15, 10. On the upper left part, we used the disc area to represent the value, while in the bottom part we used the disc radius. Results are visually very different. In the latter case (red circles), the last value (10) appears very small compared to the first one (30), while the ratio between the two values is only 3:1. This situation is actually very frequent in the literature because the command (or interface) used to produce circles or scatter plots (with varying point sizes) offers to use the radius as default to specify the disc size. It thus appears logical to use the value for the radius, but this is misleading. On the right part of the figure, we display a series of ten values using the full range for values on the top part (y axis goes from 0 to 100) or a partial range in the bottom part (y axis goes from 80 to 100), and we explicitly did not label the y-axis to enhance the confusion. The visual perception of the two series is totally different. In the top part (black series), we tend to interpret the values as very similar, while in the bottom part, we tend to believe there are significant differences. Even if we had used labels to indicate the actual range, the effect would persist because the bars are the most salient information on these figures.
doi:10.1371/journal.pcbi.1003833.g006

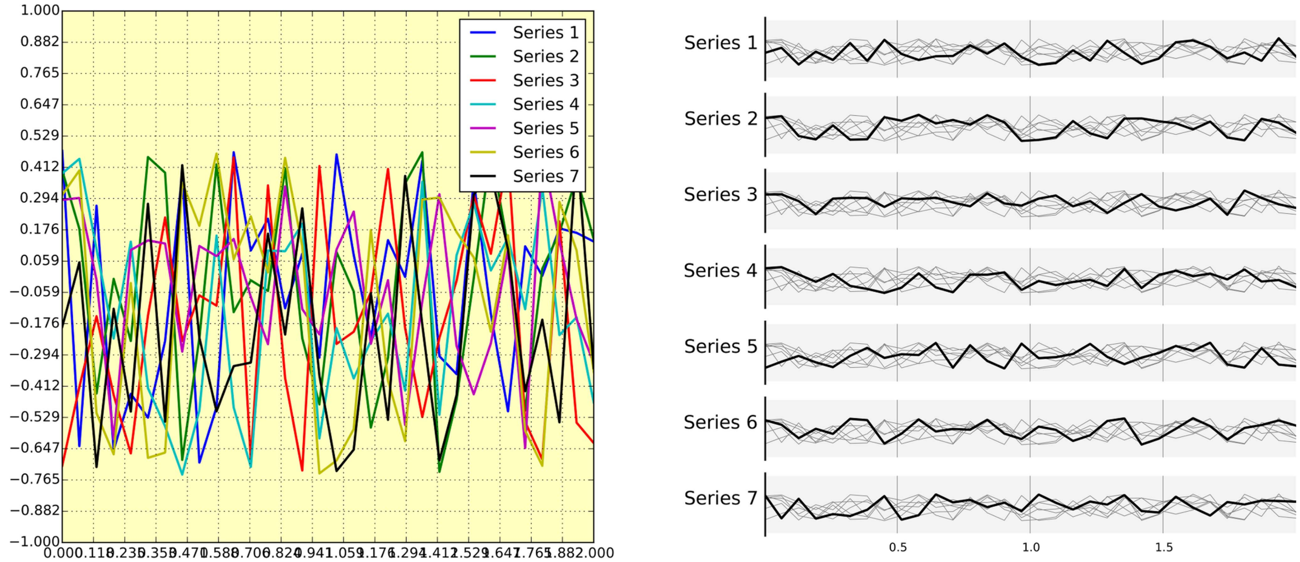



**Figure 7. Avoid chartjunk.** We have seven series of samples that are equally important, and we would like to show them all in order to visually compare them (exact signal values are supposed to be given elsewhere). The left figure demonstrates what is certainly one of the worst possible designs. All the curves cover each other and the different colors (that have been badly and automatically chosen by the software) do not help to distinguish them. The legend box overlaps part of the graphic, making it impossible to check if there is any interesting information in this area. There are far too many ticks: x labels overlap each other, making them unreadable, and the three-digit precision does not seem to carry any significant information. Finally, the grid does not help because (among other criticisms) it is not aligned with the signal, which can be considered discrete given the small number of sample points. The right figure adopts a radically different layout while using the same area on the sheet of paper. Series have been split into seven plots, each of them showing one series, while other series are drawn very lightly behind the main one. Series labels have been put on the left of each plot, avoiding the use of colors and a legend box. The number of x ticks has been reduced to three, and a thin line indicates these three values for all plots. Finally, y ticks have been completely removed and the height of the gray background boxes indicate the [−1,+1] range (this should also be indicated in the figure caption if it were to be used in an article).
doi:10.1371/journal.pcbi.1003833.g007

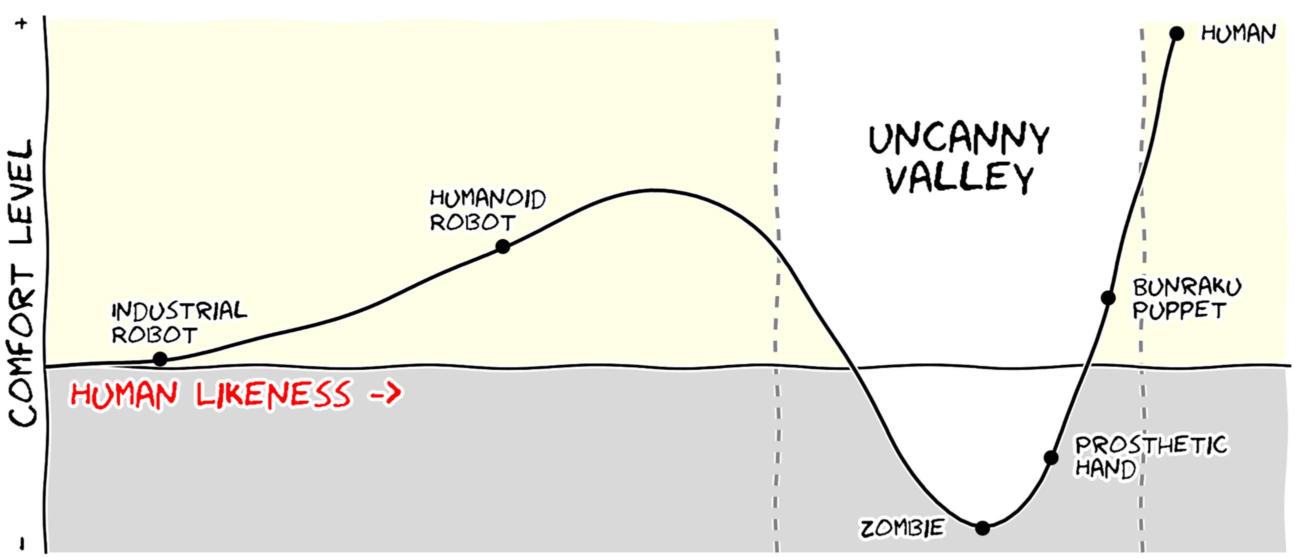



**Figure 8. Message trumps beauty.** This figure is an extreme case where the message is particularly clear even if the aesthetic of the figure is questionable. The uncanny valley is a well-known hypothesis in the field of robotics that correlates our comfort level with the human-likeness of a robot. To express this hypothetical nature, hypothetical data were used ($y = x^2 - 5e^{-5(x-2)^2}$) and the figure was given a sketched look (xkcd filter on matplotlib) associated with a cartoonish font that enhances the overall effect. Tick labels were also removed since only the overall shape of the curve matters. Using a sketch style conveys to the viewer that the data is approximate, and that it is the higher-level concepts rather than low-level details that are important [10].
doi:10.1371/journal.pcbi.1003833.g008

graphics, design, and art are becoming thinner and thinner [9]. There exists a myriad of online graphics in which aesthetic is the first criterion and content comes in second place. Even if a lot of those graphics might be considered beautiful, most of them do not fit the scientific framework. Remember, in science, message and readability of the figure is the most important aspect while beauty is only an option, as dramatically shown in Figure 8.

## Rule 10: Get the Right Tool

There exist many tools that can make your life easier when creating figures, and knowing a few of them can save you a lot of time. Depending on the type of visual you're trying to create, there is generally a dedicated tool that will do what you're trying to achieve. It is important to understand at this point that the software or library you're using to make a visualization can be different from the software or library you're using to conduct your research and/or analyze your data. You can always export data in order to use it in another tool. Whether drawing a graph, designing a schema of your experiment, or plotting some data, there are open-source tools for you. They're just waiting to be found and used. Below is a small list of open-source tools.

**Matplotlib** is a python plotting library, primarily for 2-D plotting, but with some 3-D support, which produces publication-quality figures in a variety of hardcopy formats and interactive environments across platforms. It comes with a huge gallery of examples that cover virtually all scientific domains (http://matplotlib.org/gallery.html).

**R** is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, etc.) and graphical techniques, and is highly extensible.

**Inkscape** is a professional vector graphics editor. It allows you to design complex figures and can be used, for example, to improve a script-generated figure or to read a PDF file in order to extract figures and transform them any way you like.

**TikZ and PGF** are TeX packages for creating graphics programmatically. TikZ is built on top of PGF and allows you to create sophisticated graphics in a rather intuitive and easy manner, as shown by the Tikz gallery (http://www.texample.net/tikz/examples/all/).

**GIMP** is the GNU Image Manipulation Program. It is an application for such tasks as photo retouching, image composition, and image authoring. If you need to quickly retouch an image or add some legends or labels, GIMP is the perfect tool.

**ImageMagick** is a software suite to create, edit, compose, or convert bitmap images from the command line. It can be used to quickly convert an image into another format, and the huge script gallery (http://www.fmwconcepts.com/imagemagick/index.php) by Fred Weinhaus will provide virtually any effect you might want to achieve.

**D3.js** (or just D3 for Data-Driven Documents) is a JavaScript library that offers an easy way to create and control interactive data-based graphical forms which run in web browsers, as shown in the gallery at http://github.com/mbostock/d3/wiki/Gallery.

**Cytoscape** is a software platform for visualizing complex networks and integrating these with any type of attribute data. If your data or results are very complex, cytoscape may help you alleviate this complexity.

**Circos** was originally designed for visualizing genomic data but can create figures from data in any field. Circos is useful if you have data that describes relationships or multilayered annotations of one or more scales.

## Notes

All the figures for this article were produced using matplotlib, and figure scripts are available from https://github.com/rougier/ten-rules.

## References


1. Tufte EG (1983) The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press.
2. Doumont JL (2009) Trees, maps, and theorems. Brussels: Principiae.
3. Kosara R, Mackinlay J (2013) Storytelling: The next step for visualization. IEEE Comput 46: 44–50.
4. Rougier NP (2012) Scientific visualization and matplotlib tutorial. Euroscipy 2012 & 2013. Available: http://www.loria.fr/~rougier/teaching/matplotlib/matplotlib.html. Accessed 12 August 2014.
5. Borland D, Taylor RM (2007) Rainbow color map (still) considered harmful. IEEE Comput Graph Appl 27: 14–17.
6. Okabe M, Ito K (2008). Color universal design (cud) - how to make figures and presentations that are friendly to colorblind people. Available: http://jfly.iam.u-tokyo.ac.jp/color/. Accessed 12 August 2014.
7. Few S (2011) The chartjunk debate, a close examination of recent findings. Visual Business Intelligence Newsletter. Available: http://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf. Accessed 12 August 2014.
8. Fung K (2005). Junk charts: Recycling chartjunk as junk art. Available: http://junkcharts.typepad.com. Accessed 12 August 2014.
9. Borkin MA, Vo AA, Bylinskii Z, Isola P, Sunkavalli S, et al. (2013) What makes a visualization memorable? IEEE Trans Vis Comput Graph 19: 2306–2315.
10. Schumann J, Strothotte T, Raab A, Laser S (1996) Assessing the effect of non-photorealistic rendered images in cad. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 13–18 April 1996; New York, New York, United States. CHI 96. New York: Association for Computing Machinery. pp.35–41.

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules of Live Tweeting at Scientific Conferences

**Sean Ekins[1,2]*, Ethan O. Perlstein[3]**

1 Collaborations in Chemistry, Fuquay-Varina, North Carolina, United States of America, 2 Collaborative Drug Discovery, Burlingame, California, United States of America, 3 Perlstein Lab, San Francisco, California, United States of America

The power of mobile communications has increased dramatically in recent years such that these devices (smartphone or tablet computer) can be used productively to do science [1]. The software applications installed on them do not necessarily have to be specialized to be useful for science, e.g., Evernote can be used as an electronic lab notebook [2]. Twitter is a popular microblogging platform famously limited to messages of up to 140 characters [3] and represents a simple way to express what's on your mind to a global audience of followers. Twitter has useful real-world scientific applications, such as in disease surveillance enabling the tracking of disease pandemics [4–6], as well as the capacity to be used for the communication of science itself [7]. Like other professionals, scientists are increasingly tweeting about their own research and the work of colleagues and sharing links to scholarly publications, laboratory results, and related scientific content such as molecular structures [8]. Twitter can additionally serve as a catalyst in the development of scientific tools, with at least one mobile app for science coming directly out of a tweet at a scientific conference [9].

If he or she is fortunate, a scientist may attend one or more scientific conferences in a year. In some fields, the number of conferences to attend is overwhelming. The time and cost expenditures required to physically participate in conferences necessitate an alternative route to access the information presented and capture it for future reference. Ideally, it would be preferable to monitor conferences remotely and at minimal or no cost. Increasingly, some scientists are using Twitter as a vehicle to summarize presentations and posters at conferences in real time, which is defined as "live tweeting." The advantage of remote participation is that the information tweeted is open and free to anyone around the globe (Figure 1). From our own experiences of attending and live tweeting at several conferences over the past three years, the success of live tweeting appears dependent on the engagement of conference organizers with Twitter and its active encouragement before, during, and after the meeting. Surprisingly few conferences are actively encouraging scientists to tweet. This reticence is probably more likely due to ignorance of the potential rather than the possibility of loss of attendee revenue. We suggest that conference live tweeting is an opportunity to reach beyond those in the room while enabling feedback from those outside. Obviously, it is also in the best interests of conference organizers to provide free Wi-Fi so that international attendees do not have to use their expensive data plans and because the phone signal in many conference venues is generally weak. Crucially, the success of live tweeting depends on the ability of scientists to relay the highlights of a talk or to string together multiple tweets such that they can also be read as a contiguous narrative using tools such as Storify [10]. Some simple steps to enable the wider use of live tweeting at conferences may not be widely known to scientists.

For example, conferences like "Science Online" (#scioX, in which the # is a hashtag, the keyword-tagging system of Twitter that enables retrieval of all tweets about this conference) (Box 1) are at one extreme as an "unconference" [11], with multiple vibrant discussions happening during the sessions via Twitter. These discussions extend beyond the actual physical attendees, creating a parallel virtual meeting. Live tweeting is therefore a powerful tool for expanding scientific discourse to those not fortunate enough to attend a conference in person. Similarly, if a meeting has parallel sessions, tweeting then enables conference attendees or virtual conference attendees to listen in on multiple talks simultaneously. These conferences do not have to be limited to academic gatherings and may extend to those that are organized by commercial entities, which are generally more expensive to attend and very specialized. Often useful discussions happen between talks in casual environments, and tweeting those observations or conversations is probably acceptable with the agreement of both parties, unless these are private, off-the-record discussions.

At the other extreme, which unfortunately is representative of most scientific conferences we have attended, there are few if any active live tweeters. This could be for several reasons: demographics of attendees, esoteric subject matter, and whether the organizer wants information to extend outside the conference halls (Gordon Conferences is one organization that may discourage tweeting on the assumption that this prevents scientists from sharing unpublished data). Sometimes the organizers of these conferences either do not actively encourage tweeting or they choose a cumbersome hashtag (Box 1) that consumes precious characters without signaling what the conference is even about (e.g., the Lysosomal Disease Network's #world_symposium, the annual conference on lysosomal storage diseases [12], which we shortened to

**Citation:** Ekins S, Perlstein EO (2014) Ten Simple Rules of Live Tweeting at Scientific Conferences. PLoS Comput Biol 10(8): e1003789. doi:10.1371/journal.pcbi.1003789

**Editor:** Philip E. Bourne, National Institutes of Health, United States of America

**Published** August 21, 2014



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* Email: ekinssean@yahoo.com


**Figure 1. Members of an audience at a scientific conference may tweet what they hear or see, and because these messages are free and open, it has the potential to reach anyone, anywhere in the world.** This has profound implications for the communication of science, enabling discovery, discussion, teaching, and learning outside of the confines of the conference itself. Image credit for globe: NASA Goddard Space Flight Center on Flickr.

#LDN14). Others have provided general recommendations for tweeting at academic conferences, such as rules of thumb [13,14], dos and don'ts [15], and the types of tweets that can be useful [16]. However, we are not aware of concise efforts to describe live tweeting at science conferences other than a vaguely informative ''how to tweet at conferences'' [17]. An exhaustive perspective on live coverage at scientific conferences using web technologies has been described at length and focuses on bloggers in general [18], but this does not go into detail on how to use Twitter at these conferences specifically. This is important because the types of information tweeted could also be useful to followers in different spheres, such as patients, disease advocates, financial analysts, and pharmaceutical and biotech companies.

Scientists in some cases tend to be quite introverted (varying by field) so any efforts to break the ice or engage new participants at conferences are also welcome. Twitter can play an active role here to bridge or break down the gap between researcher cliques and can serve as a means to introduce you and your ideas to others in the field, without having to personally ''know'' them. We have found from our own experiences that Twitter interactions that initially formed online during the meeting or previous meetings can have a lasting presence in real life, forging collaborations and further expanding on discussions initiated via Twitter.

In light of those observations, it's worth proposing ten simple guidelines to encourage conference organizers, conference attendees, and anyone interested who uses Twitter to enhance the spread of scientific information beyond the physical walls of the auditoria in which meetings are held. While it is possible to add many other recommendations (such as encouraging the use of Storify to combine tweets from a meeting), we believe this is a good starting point for scientists new to Twitter and perhaps previously unwilling or unable to live tweet. While we would not claim to be the absolute authorities on Twitter use at conferences, our cumulative experiences of live tweeting have enabled us to provide a short list of recommendations. These ten simple rules are certainly ripe for future refinement or replacement as other microblogging tools are developed. Of course, it's also important to remember to enjoy the conference (if you are attending in person) and please try to add some local

**Box 1. Common Twitter Abbreviations**

# = hashtag

@ = nametag, a way to reply to someone

.@ = broadcast a tweet that begins with a nametag

RT = retweet, share something already tweeted

HT = hat tip, acknowledge or thank a source

DM = direct message

CX = correction

Tweetup = physical meeting of tweeters

Additional abbreviations can be found elsewhere: http://socialmediatoday.com/emoderation/512987/top-twitter-abbreviations-you-need-know

http://www.ogawadesign.com/services/twitter-for-your-biz/twitter-abbreviations-and-twitter-acronymns.html

http://www.webopedia.com/quick_ref/Twitter_Dictionary_Guide.asp

Acronyms for common conferences can be found here: http://www.abbreviations.com/acronyms/CONF

color to the proceedings in your tweets by describing the conference locale (using pictures if permitted). Don't be afraid to add personality while providing a voice for those not physically attending.

In the style of Twitter, we have kept these "rules or recommendations" to a maximum of 140 characters (so that they can in turn be tweeted).

## Rule 1: Short Conference Hashtag

As soon as the meeting is announced, conference organizers should claim a short (6–8 characters) descriptive # that includes the year.

## Rule 2: Promote the Hashtag

Highlight the hashtag in all conference materials online, in print, on name badges, and on Twitter if possible.

## Rule 3: Encourage Tweeting

Encourage live tweeting at the conference. Session chairs can facilitate this and relay questions from the twitterosphere.

## Rule 4: Conference Twitter Etiquette

Keep questions short and on the science, avoid grandstanding, encourage responsible tweeting, and avoid harassment or snarkiness.

## Rule 5: Conference Tweet Layout

List speaker name, affiliation and conference hashtag in the first tweet; surname or initials and meeting hashtag are sufficient thereafter.

## Rule 6: Keep Conference Discussion Flowing

Summarize presentations concisely, use hashtags for keywords, and use "@ reply" to engage individuals who can add to the discussion.

## Rule 7: Differentiate Your Opinions from the Speaker's

Separate your own comments/viewpoints on the speaker or science being described in a presentation from the speaker's own words.

## Rule 8: Bring Questions up from Outside

Check for and raise questions from those outside the conference, returning the speaker responses to positively enforce participation.

## Rule 9: Meet Other Live Tweeters Face to Face

Organize tweetups so that conference attendees can meet in person and consolidate relationships and collaborations.

## Rule 10: Emphasize Impact of Live Tweeting

Ensure that positive effects of tweeting at conferences, such as discoveries, publications, or collaborations, are highlighted.


## Acknowledgments

Dr. Neil Dufton is gratefully acknowledged for creating the figure.


## References


1. Williams AJ, Ekins S, Clark AM, Jack JJ, Apodaca RL (2011) Mobile apps for chemistry in the world of drug discovery. Drug Disc Today 16: 928–939.
2. Walsh E, Cho I (2013) Using Evernote as an electronic lab notebook in a translational science laboratory. J Lab Autom 18: 229–234.
3. Anon (2014) Twitter. Available: https://twitter.com/. Accessed 14 July 2014.
4. Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS ONE 5: e14118.
5. Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS ONE 6: e19467.
6. Chunara R, Andrews JR, Brownstein JS (2012) Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 86: 39–45.
7. Ekins S, Clark AM, Williams AJ (2012) Open Drug Discovery Teams: A Chemistry Mobile App for Collaboration. Molecular Informatics 31: 585–597.
8. Ekins S, Clark AM, Williams AJ (2012) Open Drug Discovery Teams: A Chemistry Mobile App for Collaboration. Mol Inform 31: 585–597.
9. Ekins S, Clark AM, Williams AJ (2013) Incorporating Green Chemistry Concepts into Mobile Chemistry Applications and Their Potential Uses. ACS Sustain Chem Eng 1: 8–13.
10. Anon (2014) Storify. Available: https://storify.com/. Accessed 14 July 2014.
11. Anon (2014) Science Online. Available: http://scienceonline.com/. Accessed 14 July 2014.
12. Anon (2014) World Symposium. Available: http://lysosomaldiseasenetwork.org/. Accessed 14 July 2014.
13. Priego E (2012 October 3) Live-tweeting at academic conferences: 10 rules of thumb. Available: http://www.theguardian.com/higher-education-network/blog/2012/oct/03/ethics-live-tweeting-academic-conferences. Accessed 14 July 2014.
14. Croxall B (2014 January 6) Ten Tips for Tweeting at Conferences. Available: http://chronicle.com/blogs/profhacker/ten-tips-for-tweeting-at-conferences/54281. Accessed 14 July 2014.

15. Varin V (2013 March 5) The Dos and Don'ts of Live-Tweeting at an Academic Conference: An Update. Available: http://blog.historians.org/2013/03/the-dos-and-donts-of-live-tweeting-at-an-academic-conference-an-update/. Accessed 14 July 2014.
16. Long CP (2013 September 16) The art of live-tweeting. Available: http://www.cplong.org/2013/09/the-art-of-live-tweeting/. Accessed 14 July 2014.
17. Shiffman D (2012 January 17) How to live-tweet a conference: A guide for conference organizers and twitter users. Available: http://www.southernfriedscience.com/?p=12120. Accessed 14 July 2014.
18. Lister AL, Datta RS, Hofmann O, Krause R, Kuhn M, et al. (2010) Live coverage of scientific conferences using web technologies. PLoS Comput Biol 6: e1000563.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Writing Research Papers

**Weixiong Zhang***

Department of Computer Science and Engineering, Department of Genetics, Washington University in St. Louis, St. Louis, Missouri, United States of America

The importance of writing well can never be overstated for a successful professional career, and the ability to write solid papers is an essential trait of a productive researcher. Writing and publishing a paper has its own life cycle; properly following a course of action and avoiding missteps can be vital to the overall success not only of a paper but of the underlying research as well. Here, we offer ten simple rules for writing and publishing research papers.

As a caveat, this essay is not about the mechanics of composing a paper, much of which has been covered elsewhere, e.g., [1,2]. Rather, it is about the principles and attitude that can help guide the process of writing in particular and research in general. In this regard, some of the discussion will complement, extend, and refine some advice given in early articles of this Ten Simple Rules series of *PLOS Computational Biology* [3–8].

## Rule 1: Make It a Driving Force

Never separate writing a paper from the underlying research. After all, writing and research are integral parts of the overall enterprise. Therefore, design a project with an ultimate paper firmly in mind. Include an outline of the paper in the initial project design documents to help form the research objectives, determine the logical flow of the experiments, and organize the materials and data to be used. Furthermore, use writing as a tool to reassess the overall project, reevaluate the logic of the experiments, and examine the validity of the results during the research. As a result, the overall research may need to be adjusted, the project design may be revised, new methods may be devised, and new data may be collected. The process of research and writing may be repeated if necessary.

## Rule 2: Less Is More

It is often the case that more than one hypothesis or objective may be tackled in one project. It is also not uncommon that the data and results gathered for one objective can serve additional purposes. A decision on having one or more papers needs to be made, and the decision will be affected by various factors. Regardless of the validity of these factors, the overriding consideration must be the potential impact that the paper may have on the research subject and field. Therefore, the significance, completeness, and coherence of the results presented as a whole should be the principal guide for selecting the story to tell, the hypothesis to focus upon, and materials to include in the paper, as well as the yardstick for measuring the quality of the paper. By this metric, *less is more*, i.e., fewer but more significant papers serve both the research community and one's career better than more papers of less significance.

## Rule 3: Pick the Right Audience

Deciding on an angle of the story to focus upon is the next hurdle to jump at the initial stage of the writing. The results from a computational study of a biological problem can often be presented to biologists, computational scientists, or both; deciding what story to tell and from what angle to pitch the main idea is important. This issue translates to choosing a target audience, as well as an appropriate journal, to cast the main messages to. This is critical for determining the organization of the paper and the level of detail of the story, so as to write the paper with the audience in mind. Indeed, writing a paper for biologists in general is different from writing for specialists in computational biology.

## Rule 4: Be Logical

The foundation of "lively" writing for smooth reading is a sound and clear logic underlying the story of the paper. Although experiments may be carried out independently, the result from one experiment may form premises and/or provide supporting data for the next experiment. The experiments and results, therefore, must be presented in a logical order. In order to make the writing an easy process to follow, this logical flow should be determined before any other writing strategy or tactic is exercised. This logical order can also help you avoid discussing the same issue or presenting the same argument in multiple places in the paper, which may dilute the readers' attention.

An effective tactic to help develop a sound logical flow is to imaginatively create a set of figures and tables, which will ultimately be developed from experimental results, and order them in a logical way based on the information flow through the experiments. In other words, the figures and tables alone can tell the story without consulting additional material. If all or some of these figures and tables are included in the final manuscript, make every effort to make them self-contained (see Rule 5 below), a favorable feature for the paper to have. In addition, these figures and tables, as well as the threading logical flow, may be used to direct or organize research activities, reinforcing Rule 1.

## Rule 5: Be Thorough and Make It Complete

Completeness is a cornerstone for a research paper, following Rule 2. This cornerstone needs to be set in both content and presentation. First, important and relevant aspects of a hypothesis pursued in the research should be discussed with detailed supporting data. If the page limit is an issue, focus on one or two main aspects with sufficient details in the main text and leave the rest to online supporting

**Citation:** Zhang W (2014) Ten Simple Rules for Writing Research Papers. PLoS Comput Biol 10(1): e1003453. doi:10.1371/journal.pcbi.1003453

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** January 30, 2014



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: weixiong.zhang@wustl.edu

materials. As a reminder, be sure to keep the details of all experiments (e.g., parameters of the experiments and versions of software) for revision, post-publication correspondence, or importantly, reproducibility of the results. Second, don't simply state what results are presented in figures and tables, which makes the writing repetitive because they are self-contained (see below), but rather, interpret them with insights to the underlying story to be told (typically in the results section) and discuss their implication (typically in the discussion section).

Third, make the whole paper self-contained. Introduce an adequate amount of background and introductory material for the right audience (following Rule 3). A statistical test, e.g., hypergeometric tests for enrichment of a subset of objects, may be obvious to statisticians or computational biologists but may be foreign to others, so providing a sufficient amount of background is the key for delivery of the material. When an uncommon term is used, give a definition besides a reference to it. Fourth, try to avoid "making your readers do the arithmetic" [9], i.e., be clear enough so that the readers don't have to make any inference from the presented data. If such results need to be discussed, make them explicit even though they may be readily derived from other data. Fifth, figures and tables are essential components of a paper, each of which must be included for a good reason; make each of them self-contained with all required information clearly specified in the legend to guide interpretation of the data presented.

## Rule 6: Be Concise

This is a caveat to Rule 5 and is singled out to emphasize its importance. Being thorough is not a license to writing that is unnecessarily descriptive, repetitive, or lengthy. Rather, on the contrary, "simplicity is the ultimate sophistication" [10]. Overly elaborate writing is distracting and boring and places a burden on the readers. In contrast, the delivery of a message is more rigorous if the writing is precise and concise. One excellent example is Watson and Crick's Nobel-Prize-winning paper on the DNA double helix structure [11] —it is only two pages long!

## Rule 7: Be Artistic

A complete draft of a paper requires a lot of work, so it pays to go the extra mile to polish it to facilitate enjoyable reading. A paper presented as a piece of art will give referees a positive initial impression of your passion toward the research and the quality of the work, which will work in your favor in the reviewing process. Therefore, concentrate on spelling, grammar, usage, and a "lively" writing style that avoids successions of simple, boring, declarative sentences. Have an authoritative dictionary with a thesaurus and a style manual, e.g., [1], handy and use them relentlessly. Also pay attention to small details in presentation, such as paragraph indentation, page margins, and fonts. If you are not a native speaker of the language the paper is written in, make sure to have a native speaker go over the final draft to ensure correctness and accuracy of the language used.

## Rule 8: Be Your Own Judge

A complete manuscript typically requires many rounds of revision. Taking a correct attitude during revision is critical to the resolution of most problems in the writing. Be objective and honest about your work and do not exaggerate or belittle the significance of the results and the elegance of the methods developed. After working long and hard, you are an expert on the problem you studied, and *you are the best referee of your own work, after all*. Therefore, inspect the research and the paper in the context of the state of the art.

When revising a draft, purge yourself out of the picture and leave your passion for your work aside. To be concrete, put yourself completely in the shoes of a referee and scrutinize all the pieces—the significance of the work, the logic of the story, the correctness of the results and conclusions, the organization of the paper, and the presentation of the materials. In practice, you may put a draft aside for a day or two—try to forget about it completely—and then come back to it fresh, consider it as if it were someone else's writing, and read it through while trying to poke holes in the story and writing. In this process, extract the meaning literally from the language as written and do not try to use your own view to interpret or extrapolate from what was written. Don't be afraid to throw away pieces of your writing and start over from scratch if they do not pass this "not-yourself" test. This can be painful, but the final manuscript will be more logically sound and better organized.

## Rule 9: Test the Water in Your Own Backyard

It is wise to anticipate the possible questions and critiques the referees may raise and preemptively address their concerns before submission. To do so, collect feedback and critiques from others, e.g., colleagues and collaborators. Discuss your work with them and get their opinions, suggestions, and comments. A talk at a lab meeting or a departmental seminar will also help rectify potential issues that need to be addressed. If you are a graduate student, running the paper and results through the thesis committee may be effective to iron out possible problems.

## Rule 10: Build a Virtual Team of Collaborators

When a submission is rejected or poorly reviewed, don't be offended and don't take it personally. Be aware that the referees spent their time on the paper, which they might have otherwise devoted to their own research, so they are doing you a favor and helping you shape the paper to be more accessible to the targeted audience. Therefore, consider the referees as your collaborators and treat the reviews with respect. This attitude can improve the quality of your paper and research.

Read and examine the reviews objectively—the principles set in Rule 8 apply here as well. Often a criticism was raised because one of the aspects of a hypothesis was not adequately studied, or an important result from previous research was not mentioned or not consistent with yours. If a critique is about the robustness of a method used or the validity of a result, often the research needs to be redone or more data need to be collected. If you believe the referee has misunderstood a particular point, check the writing. It is often the case that improper wording or presentation misled the referee. If that's the case, revise the writing thoroughly. Don't argue without supporting data. Don't submit the paper elsewhere without additional work. This can only temporally mitigate the issue, you will not be happy with the paper in the long run, and this may hurt your reputation.

Finally, keep in mind that writing is personal, and it takes a lot of practice to find one's style. What works and what does not work vary from person to person. Undoubtedly, dedicated practice will help produce stronger papers with long-lasting impact.

## Acknowledgments

Thanks to Sharlee Climer, Richard Korf, and Kevin Zhang for critical reading of the manuscript.

## References


1. Strunk W Jr, White EB (1999) The Elements of Style. 4th edition. New York: Longman.
2. Zinsser W (2006) On Writing Well: The Classic Guide to Writing Nonfiction. 30th anniversary edition. New York: Harper Perennial.
3. Bourne PE (2005) Ten simple rules for getting published. PLOS Comput Biol 1: e57. doi:10.1371/journal.pcbi.0010057
4. Erren TC, Cullen P (2007) Ten simple rules for doing your best research, according to Hamming. PLOS Comput Biol 3: e213. doi:10.1371/journal.pcbi.0030213
5. Bourne PE (2007) Ten simple rules for making good oral presentations. PLOS Comput Biol 3: e77. doi:10.1371/journal.pcbi.0030077
6. Erren TC, Bourne PE (2007) Ten simple rules for a good poster presentation. PLOS Comput Biol 3: e102. doi: 10.1371/journal.pcbi.0030102
7. Bourne PE, Korngreen A (2006) Ten simple rules for reviewers. PLOS Comput Biol 2: e110. doi:10.1371/journal.pcbi.0020110
8. Logan DW, Sandal M, Gardner PP, Manske M, Bateman A (2010) Ten simple rules for editing Wikipedia. PLOS Comput Biol 6: e1000941. doi:10.1371/journal.pcbi.1000941
9. Johnson DS (2002) A theoretician's guide to the experimental analysis of algorithms. In Goldwasser MH, Johnson DS, McGeoch CC, editors. Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges. Providence: American Mathematical Society. pp.215–250.
10. Wikiquote page on Leonardo Da Vinci. Available: http://en.wikiquote.org/wiki/Leonardo_da_Vinci#Quotes_about_Leonardo. Accessed 13 December 2013.
11. Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. Nature 171: 737–738. Available: http://www.nature.com/nature/dna50/watsoncrick.pdf. Accessed 31 December 2013.

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Writing a Literature Review

**Marco Pautasso**[1,2]*

**1** Centre for Functional and Evolutionary Ecology (CEFE), CNRS, Montpellier, France, **2** Centre for Biodiversity Synthesis and Analysis (CESAB), FRB, Aix-en-Provence, France

Literature reviews are in great demand in most scientific fields. Their need stems from the ever-increasing output of scientific publications [1]. For example, compared to 1991, in 2008 three, eight, and forty times more papers were indexed in Web of Science on malaria, obesity, and biodiversity, respectively [2]. Given such mountains of papers, scientists cannot be expected to examine in detail every single new paper relevant to their interests [3]. Thus, it is both advantageous and necessary to rely on regular summaries of the recent literature. Although recognition for scientists mainly comes from primary research, timely literature reviews can lead to new synthetic insights and are often widely read [4]. For such summaries to be useful, however, they need to be compiled in a professional way [5].

When starting from scratch, reviewing the literature can require a titanic amount of work. That is why researchers who have spent their career working on a certain research issue are in a perfect position to review that literature. Some graduate schools are now offering courses in reviewing the literature, given that most research students start their project by producing an overview of what has already been done on their research issue [6]. However, it is likely that most scientists have not thought in detail about how to approach and carry out a literature review.

Reviewing the literature requires the ability to juggle multiple tasks, from finding and evaluating relevant material to synthesising information from various sources, from critical thinking to paraphrasing, evaluating, and citation skills [7]. In this contribution, I share ten simple rules I learned working on about 25 literature reviews as a PhD and postdoctoral student. Ideas and insights also come from discussions with coauthors and colleagues, as well as feedback from reviewers and editors.

## Rule 1: Define a Topic and Audience

How to choose which topic to review? There are so many issues in contemporary science that you could spend a lifetime of attending conferences and reading the literature just pondering what to review. On the one hand, if you take several years to choose, several other people may have had the same idea in the meantime. On the other hand, only a well-considered topic is likely to lead to a brilliant literature review [8]. The topic must at least be:

(i) interesting to you (ideally, you should have come across a series of recent papers related to your line of work that call for a critical summary),
(ii) an important aspect of the field (so that many readers will be interested in the review and there will be enough material to write it), and
(iii) a well-defined issue (otherwise you could potentially include thousands of publications, which would make the review unhelpful).

Ideas for potential reviews may come from papers providing lists of key research questions to be answered [9], but also from serendipitous moments during desultory reading and discussions. In addition to choosing your topic, you should also select a target audience. In many cases, the topic (e.g., web services in computational biology) will automatically define an audience (e.g., computational biologists), but that same topic may also be of interest to neighbouring fields (e.g., computer science, biology, etc.).

## Rule 2: Search and Re-search the Literature

After having chosen your topic and audience, start by checking the literature and downloading relevant papers. Five pieces of advice here:

(i) keep track of the search items you use (so that your search can be replicated [10]),
(ii) keep a list of papers whose pdfs you cannot access immediately (so as to retrieve them later with alternative strategies),
(iii) use a paper management system (e.g., Mendeley, Papers, Qiqqa, Sente),
(iv) define early in the process some criteria for exclusion of irrelevant papers (these criteria can then be described in the review to help define its scope), and
(v) do not just look for research papers in the area you wish to review, but also seek previous reviews.

The chances are high that someone will already have published a literature review (Figure 1), if not exactly on the issue you are planning to tackle, at least on a related topic. If there are already a few or several reviews of the literature on your issue, my advice is not to give up, but to carry on with your own literature review,

(i) discussing in your review the approaches, limitations, and conclusions of past reviews,
(ii) trying to find a new angle that has not been covered adequately in the previous reviews, and
(iii) incorporating new material that has inevitably accumulated since their appearance.

**Citation:** Pautasso M (2013) Ten Simple Rules for Writing a Literature Review. PLoS Comput Biol 9(7): e1003149. doi:10.1371/journal.pcbi.1003149

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** July 18, 2013



**Funding:** This work was funded by the French Foundation for Research on Biodiversity (FRB) through its Centre for Synthesis and Analysis of Biodiversity data (CESAB), as part of the NETSEED research project. The funders had no role in the preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: marpauta@gmail.com

When searching the literature for pertinent papers and reviews, the usual rules apply:

(i) be thorough,
(ii) use different keywords and database sources (e.g., DBLP, Google Scholar, ISI Proceedings, JSTOR Search, Medline, Scopus, Web of Science), and
(iii) look at who has cited past relevant papers and book chapters.

## Rule 3: Take Notes While Reading

If you read the papers first, and only afterwards start writing the review, you will need a very good memory to remember who wrote what, and what your impressions and associations were while reading each single paper. My advice is, while reading, to start writing down interesting pieces of information, insights about how to organize the review, and thoughts on what to write. This way, by the time you have read the literature you selected, you will already have a rough draft of the review.

Of course, this draft will still need much rewriting, restructuring, and rethinking to obtain a text with a coherent argument [11], but you will have avoided the danger posed by staring at a blank document. Be careful when taking notes to use quotation marks if you are provisionally copying verbatim from the literature. It is advisable then to reformulate such quotes with your own words in the final draft. It is important to be careful in noting the references already at this stage, so as to avoid misattributions. Using referencing software from the very beginning of your endeavour will save you time.

## Rule 4: Choose the Type of Review You Wish to Write

After having taken notes while reading the literature, you will have a rough idea of the amount of material available for the review. This is probably a good time to decide whether to go for a mini- or a full review. Some journals are now favouring the publication of rather short reviews focusing on the last few years, with a limit on the number of words and citations. A mini-review is not necessarily a minor review: it may well attract more attention from busy readers, although it will inevitably simplify some issues and leave out some relevant material due to space limitations. A full review will have the advantage of more freedom to cover in detail the complexities of a particular scientific development, but may then be left in the pile of the very important papers "to be read" by readers with little time to spare for major monographs.

There is probably a continuum between mini- and full reviews. The same point applies to the dichotomy of descriptive vs. integrative reviews. While descriptive reviews focus on the methodology, findings, and interpretation of each reviewed study, integrative reviews attempt to find common ideas and concepts from the reviewed material [12]. A similar distinction exists between narrative and systematic reviews: while narrative reviews are qualitative, systematic reviews attempt to test a hypothesis based on the published evidence, which is gathered using a

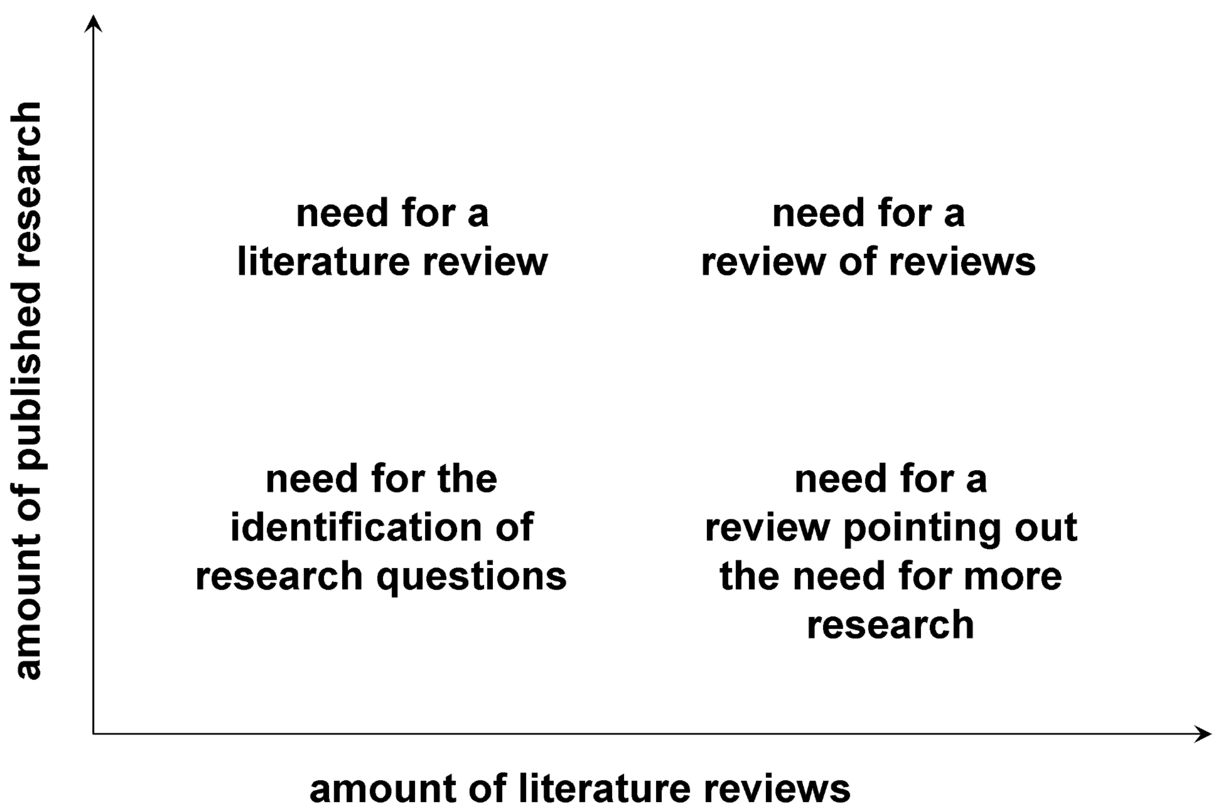



**Figure 1. A conceptual diagram of the need for different types of literature reviews depending on the amount of published research papers and literature reviews.** The bottom-right situation (many literature reviews but few research papers) is not just a theoretical situation; it applies, for example, to the study of the impacts of climate change on plant diseases, where there appear to be more literature reviews than research studies [33].
doi:10.1371/journal.pcbi.1003149.g001

predefined protocol to reduce bias [13,14]. When systematic reviews analyse quantitative results in a quantitative way, they become meta-analyses. The choice between different review types will have to be made on a case-by-case basis, depending not just on the nature of the material found and the preferences of the target journal(s), but also on the time available to write the review and the number of coauthors [15].

## Rule 5: Keep the Review Focused, but Make It of Broad Interest

Whether your plan is to write a mini- or a full review, it is good advice to keep it focused [16,17]. Including material just for the sake of it can easily lead to reviews that are trying to do too many things at once. The need to keep a review focused can be problematic for interdisciplinary reviews, where the aim is to bridge the gap between fields [18]. If you are writing a review on, for example, how epidemiological approaches are used in modelling the spread of ideas, you may be inclined to include material from both parent fields, epidemiology and the study of cultural diffusion. This may be necessary to some extent, but in this case a focused review would only deal in detail with those studies at the interface between epidemiology and the spread of ideas.

While focus is an important feature of a successful review, this requirement has to be balanced with the need to make the review relevant to a broad audience. This square may be circled by discussing the wider implications of the reviewed topic for other disciplines.

## Rule 6: Be Critical and Consistent

Reviewing the literature is not stamp collecting. A good review does not just summarize the literature, but discusses it critically, identifies methodological problems, and points out research gaps [19]. After having read a review of the literature, a reader should have a rough idea of:

(i) the major achievements in the reviewed field,
(ii) the main areas of debate, and
(iii) the outstanding research questions.

It is challenging to achieve a successful review on all these fronts. A solution can be to involve a set of complementary coauthors: some people are excellent at mapping what has been achieved, some others are very good at identifying dark clouds on the horizon, and some have instead a knack at predicting where solutions are going to come from. If your journal club has exactly this sort of team, then you should definitely write a review of the literature! In addition to critical thinking, a literature review needs consistency, for example in the choice of passive vs. active voice and present vs. past tense.

## Rule 7: Find a Logical Structure

Like a well-baked cake, a good review has a number of telling features: it is worth the reader's time, timely, systematic, well written, focused, and critical. It also needs a good structure. With reviews, the usual subdivision of research papers into introduction, methods, results, and discussion does not work or is rarely used. However, a general introduction of the context and, toward the end, a recapitulation of the main points covered and take-home messages make sense also in the case of reviews. For systematic reviews, there is a trend towards including information about how the literature was searched (database, keywords, time limits) [20].

How can you organize the flow of the main body of the review so that the reader will be drawn into and guided through it? It is generally helpful to draw a conceptual scheme of the review, e.g., with mind-mapping techniques. Such diagrams can help recognize a logical way to order and link the various sections of a review [21]. This is the case not just at the writing stage, but also for readers if the diagram is included in the review as a figure. A careful selection of diagrams and figures relevant to the reviewed topic can be very helpful to structure the text too [22].

## Rule 8: Make Use of Feedback

Reviews of the literature are normally peer-reviewed in the same way as research papers, and rightly so [23]. As a rule, incorporating feedback from reviewers greatly helps improve a review draft. Having read the review with a fresh mind, reviewers may spot inaccuracies, inconsistencies, and ambiguities that had not been noticed by the writers due to rereading the typescript too many times. It is however advisable to reread the draft one more time before submission, as a last-minute correction of typos, leaps, and muddled sentences may enable the reviewers to focus on providing advice on the content rather than the form.

Feedback is vital to writing a good review, and should be sought from a variety of colleagues, so as to obtain a diversity of views on the draft. This may lead in some cases to conflicting views on the merits of the paper, and on how to improve it, but such a situation is better than the absence of feedback. A diversity of feedback perspectives on a literature review can help identify where the consensus view stands in the landscape of the current scientific understanding of an issue [24].

## Rule 9: Include Your Own Relevant Research, but Be Objective

In many cases, reviewers of the literature will have published studies relevant to the review they are writing. This could create a conflict of interest: how can reviewers report objectively on their own work [25]? Some scientists may be overly enthusiastic about what they have published, and thus risk giving too much importance to their own findings in the review. However, bias could also occur in the other direction: some scientists may be unduly dismissive of their own achievements, so that they will tend to downplay their contribution (if any) to a field when reviewing it.

In general, a review of the literature should neither be a public relations brochure nor an exercise in competitive self-denial. If a reviewer is up to the job of producing a well-organized and methodical review, which flows well and provides a service to the readership, then it should be possible to be objective in reviewing one's own relevant findings. In reviews written by multiple authors, this may be achieved by assigning the review of the results of a coauthor to different coauthors.

## Rule 10: Be Up-to-Date, but Do Not Forget Older Studies

Given the progressive acceleration in the publication of scientific papers, today's reviews of the literature need awareness not just of the overall direction and achievements of a field of inquiry, but also of the latest studies, so as not to become out-of-date before they have been published. Ideally, a literature review should not identify as a major research gap an issue that has just been addressed in a series of papers in press (the same applies, of course, to older, overlooked studies ("sleeping beauties" [26])). This implies that literature reviewers would do well to keep an eye on electronic lists of papers in press, given that it can take months before these appear in scientific databases. Some reviews declare that they have scanned the literature up to a certain point in time, but given that peer review can be a rather lengthy process, a full search for newly appeared literature at the revision stage may be worthwhile. Assessing the contribution of papers that have just appeared is particularly challenging, because there is little perspective with which to gauge their significance and impact on further research and society.

Inevitably, new papers on the reviewed topic (including independently written literature reviews) will appear from all quarters after the review has been published, so that there may soon be the need for an updated review. But this is the nature of science [27–32]. I wish everybody good luck with writing a review of the literature.


## Acknowledgments

Many thanks to M. Barbosa, K. Dehnen-Schmutz, T. Döring, D. Fontaneto, M. Garbelotto, O. Holdenrieder, M. Jeger, D. Lonsdale, A. MacLeod, P. Mills, M. Moslonka-Lefebvre, G. Stancanelli, P. Weisberg, and X. Xu for insights and discussions, and to P. Bourne, T. Matoni, and D. Smith for helpful comments on a previous draft.


## References


1. Rapple C (2011) The role of the critical review article in alleviating information overload. Annual Reviews White Paper. Available: http://www.annualreviews.org/userimages/ContentEditor/1300384004941/Annual_Reviews_WhitePaper_Web_2011.pdf. Accessed May 2013.
2. Pautasso M (2010) Worsening file-drawer problem in the abstracts of natural, medical and social science databases. Scientometrics 85: 193–202. doi:10.1007/s11192-010-0233-5.
3. Erren TC, Cullen P, Erren M (2009) How to surf today's information tsunami: on the craft of effective reading. Med Hypotheses 73: 278–279. doi:10.1016/j.mehy.2009.05.002.
4. Hampton SE, Parker JN (2011) Collaboration and productivity in scientific synthesis. Bioscience 61: 900–910. doi:10.1525/bio.2011.61.11.9.
5. Ketcham CM, Crawford JM (2007) The impact of review articles. Lab Invest 87: 1174–1185. doi:10.1038/labinvest.3700688.
6. Boote DN, Beile P (2005) Scholars before researchers: on the centrality of the dissertation literature review in research preparation. Educ Res 34: 3–15. doi:10.3102/0013189X034006003.
7. Budgen D, Brereton P (2006) Performing systematic literature reviews in software engineering. Proc 28th Int Conf Software Engineering, ACM New York, NY, USA, pp. 1051–1052. doi:10.1145/1134285.1134500.
8. Maier HR (2013) What constitutes a good literature review and why does its quality matter? Environ Model Softw 43: 3–4. doi:10.1016/j.envsoft.2013.02.004.
9. Sutherland WJ, Fleishman E, Mascia MB, Pretty J, Rudd MA (2011) Methods for collaboratively identifying research priorities and emerging issues in science and policy. Methods Ecol Evol 2: 238–247. doi:10.1111/j.2041-210X.2010.00083.x.
10. Maggio LA, Tannery NH, Kanter SL (2011) Reproducibility of literature search reporting in medical education reviews. Acad Med 86: 1049–1054. doi:10.1097/ACM.0b013e31822221e7.
11. Torraco RJ (2005) Writing integrative literature reviews: guidelines and examples. Human Res Develop Rev 4: 356–367. doi:10.1177/1534484305278283.
12. Khoo CSG, Na JC, Jaidka K (2011) Analysis of the macro-level discourse structure of literature reviews. Online Info Rev 35: 255–271. doi:10.1108/14684521111128032.
13. Rosenfeld RM (1996) How to systematically review the medical literature. Otolaryngol Head Neck Surg 115: 53–63. doi:10.1016/S0194-5998(96)70137-7.
14. Cook DA, West CP (2012) Conducting systematic reviews in medical education: a stepwise approach. Med Educ 46: 943–952. doi:10.1111/j.1365-2923.2012.04328.x.
15. Dijkers M, The Task Force on Systematic Reviews and Guidelines (2009) The value of "traditional" reviews in the era of systematic reviewing. Am J Phys Med Rehabil 88: 423–430. doi:10.1097/PHM.0b013e31819c6c6.
16. Eco U (1977) Come si fa una tesi di laurea. Milan: Bompiani.
17. Hart C (1998) Doing a literature review: releasing the social science research imagination. London: SAGE.
18. Wagner CS, Roessner JD, Bobb K, Klein JT, Boyack KW, et al. (2011) Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. J Informetr 5: 14–26. doi:10.1016/j.joi.2010.06.004.
19. Carnwell R, Daly W (2001) Strategies for the construction of a critical review of the literature. Nurse Educ Pract 1: 57–63. doi:10.1054/nepr.2001.0008.
20. Roberts PD, Stewart GB, Pullin AS (2006) Are review articles a reliable source of evidence to support conservation and environmental management? A comparison with medicine. Biol Conserv 132: 409–423. doi:10.1016/j.biocon.2006.04.034.
21. Ridley D (2008) The literature review: a step-by-step guide for students. London: SAGE.
22. Kelleher C, Wagener T (2011) Ten guidelines for effective data visualization in scientific publications. Environ Model Softw 26: 822–827. doi:10.1016/j.envsoft.2010.12.006.
23. Oxman AD, Guyatt GH (1988) Guidelines for reading literature reviews. CMAJ 138: 697–703.
24. May RM (2011) Science as organized scepticism. Philos Trans A Math Phys Eng Sci 369: 4685–4689. doi:10.1098/rsta.2011.0177.
25. Logan DW, Sandal M, Gardner PP, Manske M, Bateman A (2010) Ten simple rules for editing Wikipedia. PLoS Comput Biol 6: e1000941. doi:10.1371/journal.pcbi.1000941.
26. van Raan AFJ (2004) Sleeping beauties in science. Scientometrics 59: 467–472. doi:10.1023/B:SCIE.0000018543.82441.f1.
27. Rosenberg D (2003) Early modern information overload. J Hist Ideas 64: 1–9. doi:10.1353/jhi.2003.0017.
28. Bastian H, Glasziou P, Chalmers I (2010) Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med 7: e1000326. doi:10.1371/journal.pmed.1000326.
29. Bertamini M, Munafò MR (2012) Bite-size science and its undesired side effects. Perspect Psychol Sci 7: 67–71. doi:10.1177/1745691611429353.
30. Pautasso M (2012) Publication growth in biological sub-fields: patterns, predictability and sustainability. Sustainability 4: 3234–3247. doi:10.3390/su4123234.
31. Michels C, Schmoch U (2013) Impact of bibliometric studies on the publication behaviour of authors. Scientometrics. doi:10.1007/s11192-013-1015-7. In press.
32. Tsafnat G, Dunn A, Glasziou P, Coiera E (2013) The automation of systematic reviews. BMJ 346: f139. doi:10.1136/bmj.f139.
33. Pautasso M, Döring TF, Garbelotto M, Pellis L, Jeger MJ (2012) Impacts of climate change on plant diseases - opinions and trends. Eur J Plant Pathol 133: 295–313. doi:10.1007/s10658-012-9936-1.

Editorial

# Ten Simple Rules for Editing Wikipedia

**Darren W. Logan[1], Massimo Sandal[2], Paul P. Gardner[1], Magnus Manske[1], Alex Bateman[1]***

1 Wellcome Trust Sanger Institute, Cambridge, United Kingdom, 2 Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

Wikipedia is the world's most successful online encyclopedia, now containing over 3.3 million English language articles. It is probably the largest collection of knowledge ever assembled, and is certainly the most widely accessible. Wikipedia can be edited by anyone with Internet access that chooses to, but does it provide reliable information? A 2005 study by *Nature* found that a selection of Wikipedia articles on scientific subjects were comparable to a professionally edited encyclopedia [1], suggesting a community of volunteers can generate and sustain surprisingly accurate content.

For better or worse, people are guided to Wikipedia when searching the Web for biomedical information [2]. So there is an increasing need for the scientific community to engage with Wikipedia to ensure that the information it contains is accurate and current. For scientists, contributing to Wikipedia is an excellent way of fulfilling public engagement responsibilities and sharing expertise. For example, some Wikipedian scientists have successfully integrated biological data with Wikipedia to promote community annotation [3,4]. This, in turn, encourages wider access to the linked data via Wikipedia. Others have used the wiki model to develop their own specialist, collaborative databases [5–8]. Taking your first steps into Wikipedia can be daunting, but here we provide some tips that should make the editing process go smoothly.

## Rule 1: Register an Account

Although any visitor can edit Wikipedia, creating a user account offers a number of benefits. Firstly, it offers you privacy and security. Though counterintuitive, editors registered under a pseudonymous username actually have greater anonymity than those who edit "anonymously". A few of us have chosen to associate our accounts with our real identities. Should you choose to forgo pseudonymity on Wikipedia, your entire editing history will be open to indefinite scrutiny by curious Web searchers, including future colleagues, students, or employers. Do not forget this.

As in academic circles, a good reputation helps your wiki career. By logging in you can build a record of good edits, and it is easier to communicate and collaborate with others if you have a fixed, reputable identity. Finally, registering an account provides access to enhanced editing features, including a "watchlist" for monitoring articles you have edited previously.

## Rule 2: Learn the Five Pillars

There are some broad principles—known as the "five pillars"—all editors are expected to adhere to when contributing to Wikipedia. Perhaps most important for scientists is the appreciation that Wikipedia is not a publisher of original thought or research [9]. Accordingly, it is not an appropriate venue to promote your pet theory or share unpublished results. It is also not a soapbox on which to expound your personal theories or a battleground to debate controversial issues. In this respect, Wikipedia fundamentally differs from other types of new media, such as blogs, that encourage editorializing.

Contributing to Wikipedia is something to enjoy; a natural extension of your enthusiasm for science. But differences of opinion inevitably arise, particularly on pages provided for discussion on how to improve articles. Treat other editors as collaborators and maintain a respectful and civil manner, even in disagreement [10]. If you begin to find a particular interaction stressful, simply log off and come back another time. Unlike most scientific enterprises, Wikipedia has no deadlines.

## Rule 3: Be Bold, but Not Reckless

The survival and growth of any wiki requires participation. Wikipedia is unmatched in size, but its continuing success depends on the regular contributions of tens of thousands of volunteers. Therefore, Wikipedia urges all users to be bold: if you spot an error, correct it. If you can improve an article, please do so. It is important, however, to distinguish boldness from recklessness. Start off small. Begin by making minor modifications to existing articles before attempting a complete rewrite of *History of science*.

Many new editors feel intimidated about contributing to Wikipedia at first, fearing they may a mistake. Such reticence is understandable but unfounded. The worst that can happen is your first edits are deemed not to be an improvement and they get reverted. If this does occur, treat it as a positive learning experience and ask the reverting editor for advice.

## Rule 4: Know Your Audience

Wikipedia is not primarily aimed at experts; therefore, the level of technical detail in its articles must be balanced against the ability of non-experts to understand those details. When contributing scientific content, imagine you have been tasked with writing a comprehensive scientific review for a high school audience. It can be surprisingly challenging explaining complex ideas in an accessible, jargon-free manner. But it is worth the perseverance. You will reap the benefits when it comes to writing your next manuscript or teaching an undergraduate class.

## Rule 5: Do Not Infringe Copyright

With certain conditions, almost all of Wikipedia's content is free for anyone to reuse, adapt, and distribute. Consequently,

**Citation:** Logan DW, Sandal M, Gardner PP, Manske M, Bateman A (2010) Ten Simple Rules for Editing Wikipedia. PLoS Comput Biol 6(9): e1000941. doi:10.1371/journal.pcbi.1000941

**Published** September 30, 2010



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: agb@sanger.ac.uk

it does not accept non-free material under copyright restriction. Some journals, including those from the Public Library of Science, publish material under an open-access license that is compatible with use in Wikipedia if properly attributed. Most do not. Therefore, although it may be tempting, avoid copying text or figures from your latest review article (or anyone else's) into Wikipedia. It will quickly be identified as a copyright violation and flagged for immediate deletion.

You can give Wikipedia permission to use material you own, but this process is non-reversible and can be time consuming. It is often better to rewrite the text in simpler language or redraw the figure to make it more accessible. This will also ensure it is more suitable for Wikipedia's non-expert readership (see Rule 4).

## Rule 6: Cite, Cite, Cite

To maintain the highest standards possible, Wikipedia has a strict inclusion policy that demands verifiability [11]. This is best established by attributing each statement in Wikipedia to a reliable, published source (but see Rules 7 and 8 on excessive self-citing). Most scientists are in the fortunate position of having access to a wide body of literature, and experience in using inline citations to support their writing. Since unverified content may be removed from Wikipedia at any time, provide supporting citations for every statement that might be challenged by another editor at some point in the future. Whenever possible, give preference to secondary sources (such as reviews or book chapters) that survey the relevant primary research over research articles themselves.

Wikipedia's accessibility makes each of its scientific articles an excellent entry point for laypeople seeking specialist information. By also providing direct hyperlinks to reliable, freely accessible online resources with your citations (biological databases or open-access journals, for example), other editors can quickly verify your content and readers have immediate access to authoritative sources that address the subject in greater detail.

## Rule 7: Avoid Shameless Self-Promotion

Many people are tempted to write or edit Wikipedia articles about themselves. Resist that urge. If you are sufficiently notable to merit inclusion in an encyclopedia, eventually someone else will write an article about you. Remember that unlike a personal Web page, your Wikipedia biography is not yours to control. A lovingly crafted hagiography extolling your many virtues can rapidly accumulate information you would rather not be publicized. You may already have a Wikipedia biography, but it contains factual inaccuracies that you wish to correct. How do you do this without breaking the rules? Wikipedia's guidelines encourage you to provide information about yourself on the associated discussion page, but please permit other editors to add it to the article itself.

Think twice, also, before writing about your mentors, colleagues, competitors, inventions, or projects. Doing so places you in a conflict of interest and inclines you towards unintentional bias [12]. If you have a personal or financial interest in the subject of *any* article you choose to edit, declare it on the associated discussion page and heed the advice of other editors who can offer a more objective perspective.

## Rule 8: Share Your Expertise, but Don't Argue from Authority

Writing about a subject about which you have academic expertise is not a conflict of interest [12]; indeed, this is where we can contribute to Wikipedia most effectively. Jimmy Wales, co-founder of Wikipedia, told *Nature* that experts have the ability to "write specifics in a nuanced way", thereby significantly improving article quality [1]. When writing in your area of expertise, referencing material you have published in peer-reviewed journals is permitted if it is genuinely notable, but use common sense (and revisit Rule 7). For example, if you have an obscure, never-been-cited article in the *Journal of New Zealand Dairy Research* discussing the RNA content of cow milk, then referencing this in the introductory paragraph of the Wikipedia articles on "RNA", "Milk", "Cow", and "Evolution of mammals" is not a good idea.

Occasionally you may interact with another editor who clearly does not share your expertise on the subject of an article. This can often prove frustrating for experts and is the basis of much academic angst on Wikipedia [1]. On such occasions, remember that you are assessed only on your contributions to Wikipedia, not who you are, your qualifications, or what you have achieved in your career. Your specialist knowledge should enable you to write in a neutral manner and produce reliable, independent sources to support each assertion you make. If you do not provide verification, your contributions will be rightly challenged irrespective of how many degrees you hold.

## Rule 9: Write Neutrally and with Due Weight

All articles in Wikipedia should be impartial in tone and content [13]. When writing, do state facts and facts about notable opinions, but do not offer *your* opinion as fact. Many newcomers to Wikipedia gravitate to articles on controversial issues about which people hold strong opposing viewpoints. Avoid these until familiar with Wikipedia's policies (see Rule 3), and instead focus on articles that are much easier to remain dispassionate about.

Many scientists who contribute to Wikipedia fail to appreciate that a neutral point of view is not the same as the mainstream scientific point of view. When writing about complex issues, try to cover all significant viewpoints and afford each with due weight, but not equal weight. For example, an article on a scientific controversy should describe both the scientific consensus and significant fringe theories, but not in the same depth or in a manner suggesting these viewpoints are equally held.

## Rule 10: Ask for Help

Wikipedia can be a confusing place for the inexperienced editor. Learning Wiki markup—the syntax that instructs the software how to render the page—may appear daunting at first, though the recent implementation of a new editing toolbar has made this easier, and usability development is ongoing. The intersecting guidelines and policies (and the annoying tendency of experienced editors to use an alphabet soup of acronyms to reference them) can also be tricky to comprehend. Thankfully, the Wikipedia community puts great stock in welcoming new editors. Guidance is available through a number of avenues, including help desks, a specific IRC channel, and an Adopt-a-User mentorship program. You can even summon help using a special template—{{helpme}}—and, as if by magic, a friendly Wikipedian will appear to offer one-on-one assistance.

## Acknowledgments

We thank Philip Bourne for insightful comments and gratefully acknowledge the contributions of the many Wikipedians who collectively developed the policies and guidelines that inspired these rules.

## References


1. Giles J (2005) Internet encyclopaedias go head to head. Nature 438: 900–901.
2. Laurent MR, Vickers TJ (2009) Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc 16: 471–479.
3. Daub J, Gardner PP, Tate J, Ramskold D, Manske M, et al. (2008) The RNA WikiProject: community annotation of RNA families. RNA 14: 2462–2464.
4. Huss JW, 3rd, Orozco C, Goodale J, Wu C, Batalov S, et al. (2008) A gene wiki for community annotation of gene function. PLoS Biol 6: e175. doi:10.1371/journal.pbio.0060175.
5. Hoffmann R (2008) A wiki for the life sciences where authorship matters. Nat Genet 40: 1047–1051.
6. Mons B, Ashburner M, Chichester C, van Mulligen E, Weeber M, et al. (2008) Calling on a million minds for community annotation in WikiProteins. Genome Biol 9: R89.
7. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, et al. (2008) WikiPathways: pathway editing for the people. PLoS Biol 6: e184. doi:10.1371/journal.pbio.0060184.
8. Hodis E, Prilusky J, Martz E, Silman I, Moult J, et al. (2008) Proteopedia - a scientific 'wiki' bridging the rift between three-dimensional structure and function of biomacromolecules. Genome Biol 9: R121.
9. Wikipedia contributors (2010) No original research. Wikipedia, the Free Encyclopedia. Available: http://en.wikipedia.org/wiki/Wikipedia: ><?show=to]No_original_research. Accessed 26 July 2010.
10. Wikipedia contributors (2010) Civility. Wikipedia, the Free Encyclopedia. Available: http://en.wikipedia.org/wiki/Wikipedia:Civility. Accessed 26 July 2010.
11. Wikipedia contributors (2010) Verifiability. Wikipedia, the Free Encyclopedia. Available: http://en.wikipedia.org/wiki/Wikipedia:Verifiability. Accessed 26 July 2010.
12. Wikipedia contributors (2010) Conflict of interest. Wikipedia, the Free Encyclopedia. Available: http://en.wikipedia.org/wiki/Wikipedia:Conflict_of_interest. Accessed 26 July 2010.
13. Wikipedia contributors (2010) Neutral point of view. Wikipedia, the Free Encyclopedia. Available: http://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view. Accessed 26 July 2010.

Editorial

# Ten Simple Rules for a Good Poster Presentation

**Thomas C. Erren*, Philip E. Bourne**

Posters are a key component of communicating your science and an important element in a successful scientific career. Posters, while delivering the same high-quality science, offer a different medium from either oral presentations [1] or published papers [2], and should be treated accordingly. Posters should be considered a snapshot of your work intended to engage colleagues in a dialog about the work, or, if you are not present, to be a summary that will encourage the reader to want to learn more. Many a lifelong collaboration [3] has begun in front of a poster board. Here are ten simple rules for maximizing the return on the time-consuming process of preparing and presenting an effective poster.

## Rule 1: Define the Purpose

The purpose will vary depending on the status and nature of the work being presented, as well as the intent. Some posters are designed to be used again and again; for example, those making conference attendees aware of a shared resource. Others will likely be used once at a conference and then be relegated to the wall in the laboratory. Before you start preparing the poster, ask yourself the following questions: What do you want the person passing by your poster to do? Engage in a discussion about the content? Learn enough to go off and want to try something for themselves? Want to collaborate? All the above, or none of the above but something else? Style your poster accordingly.

## Rule 2: Sell Your Work in Ten Seconds

Some conferences will present hundreds of posters; you will need to fight for attention. The first impressions of your poster, and to a lesser extent what you might say when standing in front of it, are crucial. It is analogous to being in an elevator and having a few seconds to peak someone's interest before they get off. The sad truth is that you have to sell your work. One approach is to pose your work as addressing a decisive question, which you then address as best you can. Once you have posed the question, which may well also be the motivation for the study, the focus of your poster should be on addressing that question in a clear and concise way.

## Rule 3: The Title Is Important

The title is a good way to sell your work. It may be the only thing the conference attendee sees before they reach your poster. The title should make them want to come and visit. The title might pose a decisive question, define the scope of the study, or hint at a new finding. Above all, the title should be short and comprehensible to a broad audience. The title is your equivalent of a newspaper headline—short, sharp, and compelling.

## Rule 4: Poster Acceptance Means Nothing

Do not take the acceptance of a poster as an endorsement of your work. Conferences need attendees to be financially viable. Many attendees who are there on grants cannot justify attending a conference unless they present. There are a small number of speaking slots compared with attendees. How to solve the dilemma? Enter posters; this way everyone can present. In other words, your poster has not been endorsed, just accepted. To get endorsement from your peers, do good science and present it well on the poster.

## Rule 5: Many of the Rules for Writing a Good Paper Apply to Posters, Too

Identify your audience and provide the appropriate scope and depth of content. If the conference includes nonspecialists, cater to them. Just as the abstract of a paper needs to be a succinct summary of the motivation, hypothesis to be tested, major results, and conclusions, so does your poster.

## Rule 6: Good Posters Have Unique Features Not Pertinent to Papers

The amount of material presented in a paper far outweighs what is presented on a poster. A poster requires you to distill the work, yet not lose the message or the logical flow. Posters need to be viewed from a distance, but can take advantage of your presence. Posters can be used as a distribution medium for copies of associated papers, supplementary information, and other handouts. Posters allow you to be more speculative. Often only the titles or at most the abstracts of posters can be considered published; that is, widely distributed. Mostly, they may never be seen again. There is the opportunity to say more than you would in the traditional literature, which for all intents and purposes will be part of the immutable record. Take advantage of these unique features.

## Rule 7: Layout and Format Are Critical

Pop musician Keith Richards put the matter well in an interview with *Der Spiegel* [4]: "If you are a painter, then the most important thing is the bare canvas. A good painter will never cover all the space but will always leave some

**Citation:** Erren TC, Bourne PE (2007) Ten simple rules for a good poster presentation. PLoS Comput Biol 3(5): e102. doi:10.1371/journal.pcbi.0030102



Thomas C. Erren is with the Institute and Policlinic for Occupational and Social Medicine, School of Medicine and Dentistry, University of Cologne, Lindenthal, Germany. Philip E. Bourne is a Professor in the Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America.

* To whom correspondence should be addressed. E-mail: tim.erren@uni-koeln.de

blank. My canvas is silence." Your canvas as poster presenter is also white space. Guide the passerby's eyes from one succinct frame to another in a logical fashion from beginning to end. Unlike the literature, which is linear by virtue of one page following another, the reader of a poster is free to wander over the pages as if they are tacked to the poster board in a random order. Guide the reader with arrows, numbering, or whatever else makes sense in getting them to move from one logical step to another. Try to do this guiding in an unusual and eye-catching way. Look for appropriate layouts in the posters of others and adopt some of their approaches. Finally, never use less than a size 24 point font, and make sure the main points can be read at eye level.

### Rule 8: Content Is Important, but Keep It Concise

Everything on the poster should help convey the message. The text must conform to the norms of sound scientific reporting: clarity, precision of expression, and economy of words. The latter is particularly important for posters because of their inherent space limitations. Use of first-rate pictorial material to illustrate a poster can sometimes transform what would otherwise be a bewildering mass of complex data into a coherent and convincing story. One carefully produced chart or graph often says more than hundreds of words. Use graphics for "clear portrayal of complexity" [5], not to impress (and possibly bewilder) viewers with complex artistry. Allow a figure to be viewed in both a superficial and a detailed way. For example, a large table might have bold swaths of color indicating relative contributions from different categories, and the smaller text in the table would provide gritty details for those who want them. Likewise, a graph could provide a bold trend line (with its interpretation clearly and concisely stated), and also have many detailed points with error bars. Have a clear and obvious set of conclusions—after the abstract, this is where the passerby's eyes will wander. Only then will they go to the results, followed by the methods.

### Rule 9: Posters Should Have Your Personality

A poster is a different medium from a paper, which is conventionally dry and impersonal. Think of your poster as an extension of your personality. Use it to draw the passerby to take a closer look or to want to talk to you. Scientific collaboration often starts for reasons other than the shared scientific interest, such as a personal interest. A photo of you on the poster not only helps someone find you at the conference when you are not at the poster, it can also be used to illustrate a hobby or an interest that can open a conversation.

### Rule 10: The Impact of a Poster Happens Both During and After the Poster Session

When the considerable effort of making a poster is done, do not blow it on presentation day by failing to have the poster achieve maximum impact. This requires the right presenter–audience interaction. Work to get a crowd by being engaging; one engaged viewer will attract others. Don't badger people, let them read. Be ready with Rule 2. Work all the audience at once, do not leave visitors waiting for your attention. Make eye contact with every visitor.

Make it easy for a conference attendee to contact you afterward. Have copies of relevant papers on hand as well as copies of the poster on standard-sized paper. For work that is more mature, have the poster online and make the URL available as a handout. Have your e-mail and other demographics clearly displayed. Follow up with people who come to the poster by having a signup sheet.

The visitor is more likely to remember you than the content of your poster. Make yourself easy to remember. As the host of the work presented on the poster, be attentive, open, and curious, and self-confident but never arrogant and aggressive. Leave the visitors space and time—they can "travel" through your poster at their own discretion and pace. If a visitor asks a question, talk simply and openly about the work. This is likely your opportunity to get feedback on the work before it goes to publication. Better to be tripped up in front of your poster than by a reviewer of the manuscript.

Good posters and their presentations can improve your reputation, both within and outside your working group and institution, and may also contribute to a certain scientific freedom. Poster prizes count when peers look at your resume.

These ten rules will hopefully help you in preparing better posters. For a more humorous view on what not to do in preparing a poster, see [6], and for further information, including the opportunity to practice your German, see [7]. ■


## Acknowledgments

Thomas Erren's contributions to this piece are based on [7] and were stimulated by exchanges with Michael Jacobsen. Thanks also to Steven E. Brenner for useful input.

**Funding.** The authors received no specific funding for this article.

**Competing interests.** The authors have declared that no competing interests exist.



**References**

1. Bourne PE (2007) Ten simple rules for making good oral presentations. PLoS Comput Biol 3: e77. doi:10.1371/journal.pcbi.0030077
2. Bourne PE (2005) Ten simple rules for getting published. PLoS Comput Biol 1: e57. doi:10.1371/journal.pcbi.0010057
3. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comput Biol 3: e44. doi:10.1371/journal.pcbi.0030044
4. (1998) Interview with Keith Richards. Meine Leinwand ist die Stille. Der Spiegel 45: 167–170.
5. Tufte ER (2001) The visual display of quantitative information. Cheshire (Connecticut): Graphics Press. p. 191.
6. Wolcott TG (1997) Mortal sins in poster presentations or how to give the poster no one remembers. Newsletter Soc Integr Compar Biol Fall: 10–11. Available: http://www.sicb.org/newsletters/fa97nl/sicb/poster.html. Accessed 23 April 2007.
7. Erren TC (2006). Schau mich an! Ein Leitfaden zur Erstellung und Präsentation von Postern in der Medizin und den Naturwissenschaften. München/Wien/New York: W. Zuckschwerdt Verlag.

Editorial

# Ten Simple Rules for Making Good Oral Presentations

**Philip E. Bourne**

Continuing our "Ten Simple Rules" series [1–5], we consider here what it takes to make a good oral presentation. While the rules apply broadly across disciplines, they are certainly important from the perspective of this readership. Clear and logical delivery of your ideas and scientific results is an important component of a successful scientific career. Presentations encourage broader dissemination of your work and highlight work that may not receive attention in written form.

## Rule 1: Talk to the Audience

We do not mean face the audience, although gaining eye contact with as many people as possible when you present is important since it adds a level of intimacy and comfort to the presentation. We mean prepare presentations that address the target audience. Be sure you know who your audience is—what are their backgrounds and knowledge level of the material you are presenting and what they are hoping to get out of the presentation? Off-topic presentations are usually boring and will not endear you to the audience. Deliver what the audience wants to hear.

## Rule 2: Less is More

A common mistake of inexperienced presenters is to try to say too much. They feel the need to prove themselves by proving to the audience that they know a lot. As a result, the main message is often lost, and valuable question time is usually curtailed. Your knowledge of the subject is best expressed through a clear and concise presentation that is provocative and leads to a dialog during the question-and-answer session when the audience becomes active participants. At that point, your knowledge of the material will likely become clear. If you do not get any questions, then you have not been following the other rules. Most likely, your presentation was either incomprehensible or trite. A side effect of too much material is that you talk too quickly, another ingredient of a lost message.

## Rule 3: Only Talk When You Have Something to Say

Do not be overzealous about what you think you will have available to present when the time comes. Research never goes as fast as you would like. Remember the audience's time is precious and should not be abused by presentation of uninteresting preliminary material.

## Rule 4: Make the Take-Home Message Persistent

A good rule of thumb would seem to be that if you ask a member of the audience a week later about your presentation, they should be able to remember three points. If these are the key points you were trying to get across, you have done a good job. If they can remember any three points, but not the key points, then your emphasis was wrong. It is obvious what it means if they cannot recall three points!

## Rule 5: Be Logical

Think of the presentation as a story. There is a logical flow—a clear beginning, middle, and an end. You set the stage (beginning), you tell the story (middle), and you have a big finish (the end) where the take-home message is clearly understood.

## Rule 6: Treat the Floor as a Stage

Presentations should be entertaining, but do not overdo it and do know your limits. If you are not humorous by nature, do not try and be humorous. If you are not good at telling anecdotes, do not try and tell anecdotes, and so on. A good entertainer will captivate the audience and increase the likelihood of obeying Rule 4.

## Rule 7: Practice and Time Your Presentation

This is particularly important for inexperienced presenters. Even more important, when you give the presentation, stick to what you practice. It is common to deviate, and even worse to start presenting material that you know less about than the audience does. The more you practice, the less likely you will be to go off on tangents. Visual cues help here. The more presentations you give, the better you are going to get. In a scientific environment, take every opportunity to do journal club and become a teaching assistant if it allows you to present. An important talk should not be given for the first time to an audience of peers. You should have delivered it to your research collaborators who will be kinder and gentler but still point out obvious discrepancies. Laboratory group meetings are a fine forum for this.

## Rule 8: Use Visuals Sparingly but Effectively

Presenters have different styles of presenting. Some can captivate the audience with no visuals (rare); others require visual cues and in addition, depending on the material, may not be able to present a particular topic well without the appropriate visuals such as graphs and charts. Preparing good visual materials will be the subject of a further Ten Simple Rules. Rule 7 will

**Citation:** Bourne PE (2007) Ten simple rules for making good oral presentations. PLoS Comput Biol 3(4): e77. doi:10.1371/journal.pcbi.0030077



Dr. Philip E. Bourne is a Professor in the Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America. E-mail: bourne@sdsc.edu

help you to define the right number of visuals for a particular presentation. A useful rule of thumb for us is if you have more than one visual for each minute you are talking, you have too many and you will run over time. Obviously some visuals are quick, others take time to get the message across; again Rule 7 will help. Avoid reading the visual unless you wish to emphasize the point explicitly, the audience can read, too! The visual should support what you are saying either for emphasis or with data to prove the verbal point. Finally, do not overload the visual. Make the points few and clear.

### Rule 9: Review Audio and/or Video of Your Presentations

There is nothing more effective than listening to, or listening to and viewing, a presentation you have made. Violations of the other rules will become obvious. Seeing what is wrong is easy, correcting it the next time around is not. You will likely need to break bad habits that lead to the violation of the other rules. Work hard on breaking bad habits; it is important.

### Rule 10: Provide Appropriate Acknowledgments

People love to be acknowledged for their contributions. Having many gratuitous acknowledgements degrades the people who actually contributed. If you defy Rule 7, then you will not be able to acknowledge people and organizations appropriately, as you will run out of time. It is often appropriate to acknowledge people at the beginning or at the point of their contribution so that their contributions are very clear.

As a final word of caution, we have found that even in following the Ten Simple Rules (or perhaps thinking we are following them), the outcome of a presentation is not always guaranteed. Audience–presenter dynamics are hard to predict even though the metric of depth and intensity of questions and off-line followup provide excellent indicators. Sometimes you are sure a presentation will go well, and afterward you feel it did not go well. Other times you dread what the audience will think, and you come away pleased as punch. Such is life. As always, we welcome your comments on these Ten Simple Rules by Reader Response. ■


## Acknowledgments

The idea for this particular Ten Simple Rules was inspired by a conversation with Fiona Addison.

**Funding.** The author received no specific funding for this article.

**Competing interests.** The author has declared that no competing interests exist.



**References**

1. Bourne PE (2005) Ten simple rules for getting published. PLoS Comp Biol 1: e57.
2. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comp Biol 2: e12.
3. Bourne PE, Korngreen A (2006) Ten simple rules for reviewers. PLoS Comp Biol 2: e110.
4. Bourne PE, Friedberg I (2006) Ten simple rules for selecting a postdoctoral fellowship. PLoS Comp Biol 2: e121.
5. Vicens Q, Bourne PE (2007) Ten simple rules for a successful collaboration. PLoS Comp Biol 3: e44.


Editorial

# Ten Simple Rules for Getting Published

Philip E. Bourne

The student council (http://www.iscbsc.org/) of the International Society for Computational Biology asked me to present my thoughts on getting published in the field of computational biology at the Intelligent Systems in Molecular Biology conference held in Detroit in late June of 2005. Close to 200 bright young souls (and a few not so young) crammed into a small room for what proved to be a wonderful interchange among a group of whom approximately one-half had yet to publish their first paper. The advice I gave that day I have modified and present as ten rules for getting published.

## Rule 1: Read many papers, and learn from both the good and the bad work of others.

It is never too early to become a critic. Journal clubs, where you critique a paper as a group, are excellent for having this kind of dialogue. Reading at least two papers a day in detail (not just in your area of research) and thinking about their quality will also help. Being well read has another potential major benefit—it facilitates a more objective view of one's own work. It is too easy after many late nights spent in front of a computer screen and/or laboratory bench to convince yourself that your work is the best invention since sliced bread. More than likely it is not, and your mentor is prone to falling into the same trap, hence rule 2.

## Rule 2: The more objective you can be about your work, the better that work will ultimately become.

Alas, some scientists will never be objective about their own work, and will never make the best scientists—learn objectivity early, the editors and reviewers have.

## Rule 3: Good editors and reviewers will be objective about your work.

The quality of the editorial board is an early indicator of the review process. Look at the masthead of the journal in which you plan to publish. Outstanding editors demand and get outstanding reviews. Put your energy into improving the quality of the manuscript *before submission*. Ideally, the reviews will improve your paper. But they will not get to imparting that advice if there are fundamental flaws.

## Rule 4: If you do not write well in the English language, take lessons early; it will be invaluable later.

This is not just about grammar, but more importantly comprehension. The best papers are those in which complex ideas are expressed in a way that those who are less than immersed in the field can understand. Have you noticed that the most renowned scientists often give the most logical and simply stated yet stimulating lectures? This extends to their written work as well. Note that writing clearly is valuable, even if your ultimate career does not hinge on producing good scientific papers in English language journals. Submitted papers that are not clearly written in good English, unless the science is truly outstanding, are often rejected or at best slow to publish since they require extensive copyediting.

## Rule 5: Learn to live with rejection.

A failure to be objective can make rejection harder to take, and you will be rejected. Scientific careers are full of rejection, even for the best scientists. The correct response to a paper being rejected or requiring major revision is to listen to the reviewers and respond in an objective, not subjective, manner. Reviews reflect how your paper is being judged—learn to live with it. If reviewers are unanimous about the poor quality of the paper, move on—in virtually all cases, they are right. If they request a major revision, do it and address every point they raise both in your cover letter and through obvious revisions to the text. Multiple rounds of revision are painful for all those concerned and slow the publishing process.

## Rule 6: The ingredients of good science are obvious—novelty of research topic, comprehensive coverage of the relevant literature, good data, good analysis including strong statistical support, and a thought-provoking discussion. The ingredients of good science reporting are obvious—good organization, the appropriate use of tables and figures, the right length, writing to the intended audience—do not ignore the obvious.

Be objective about these ingredients when you review the first draft, and do not rely on your mentor. Get a candid opinion by having the paper read by colleagues without a vested interest in the work, including those not directly involved in the topic area.

## Rule 7: Start writing the paper the day you have the idea of what questions to pursue.

Some would argue that this places too much emphasis on publishing, but it could also be argued that it helps define scope and facilitates hypothesis-driven science. The temptation of novice authors is to try to include everything they know in a paper. Your thesis is/was your kitchen sink. Your papers should be concise, and impart as much information as possible in the least number of words. Be familiar with the guide to authors and follow it, the editors and reviewers do. Maintain a good bibliographic database as you go, and read the papers in it.

Citation: Bourne PE (2005) Ten simple rules for getting published. PLoS Comput Biol 1(5): e57.



DOI: 10.1371/journal.pcbi.0010057

Philip E. Bourne is Editor-in-Chief of *PLoS Computational Biology*. E-mail: bourne@sdsc.edu

### Rule 8: Become a reviewer early in your career.

Reviewing other papers will help you write better papers. To start, work with your mentors; have them give you papers they are reviewing and do the first cut at the review (most mentors will be happy to do this). Then, go through the final review that gets sent in by your mentor, and where allowed, as is true of this journal, look at the reviews others have written. This will provide an important perspective on the quality of your reviews and, hopefully, allow you to see your own work in a more objective way. You will also come to understand the review process and the quality of reviews, which is an important ingredient in deciding where to send your paper.

### Rule 9: Decide early on where to try to publish your paper.

This will define the form and level of detail and assumed novelty of the work you are doing. Many journals have a presubmission enquiry system available—use it. Even before the paper is written, get a sense of the novelty of the work, and whether a specific journal will be interested.

### Rule 10: Quality is everything.

It is better to publish one paper in a quality journal than multiple papers in lesser journals. Increasingly, it is harder to hide the impact of your papers; tools like Google Scholar and the ISI Web of Science are being used by tenure committees and employers to define metrics for the quality of your work. It used to be that just the journal name was used as a metric. In the digital world, everyone knows if a paper has little impact. Try to publish in journals that have high impact factors; chances are your paper will have high impact, too, if accepted.

When you are long gone, your scientific legacy is, in large part, the literature you left behind and the impact it represents. I hope these ten simple rules can help you leave behind something future generations of scientists will admire. ■

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for developing good reading habits during graduate school and beyond

**Marcos Méndez** *

Area of Biodiversity and Conservation, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

* marcos.mendez@urjc.es

## Introduction

Scientific activity has been compared to a scholarly conversation [1]. As scientific communities are now very large, this conversation mostly proceeds by writing and reading. While lot of advice exists about the writing part [2–4], much less attention has been devoted to what to read (see, however, [5]). Reading is fundamental for keeping updated in the advances of your discipline. However, scholarly literature grows at an increasingly fast pace [6]. Scientists, regardless of the stage in their career, have trouble in navigating an overwhelming amount of relevant literature. This overload can potentially lead to increasingly selective or cursory reading. Consequences are detrimental for the training of new scientists by decreasing quality standards. Good reading habits are essential and need to be instilled from the very beginning of a scientific career. As a coordinator of a doctoral program, I have addressed this issue in a recent workshop, after realizing the demand of advice on this topic by many graduate students. Here, I summarize the content of this workshop in 10 simple rules for achieving good reading habits (see also [5]). They were mainly intended to guide early graduate students, but I believe they can also be useful at later career stages.

## Rule 1: Develop the habit of reading on a daily basis

A first step toward good reading habits is to realize that reading is a fundamental part of your training (or activity) as a researcher. If you do not read regularly, you will soon get out of date and will not be able to join the scholarly conversation. Piling or archiving unread papers (see Rule 9) only leads to a delusion of knowledge. Cursory or urgent reading when preparing a new article is neither efficient nor desirable. Instead, dedicate daily time slot for reading at least one paper. When to read is a very personal decision. I use my daily commuting time in public transport for that task; others start their working day doing some reading. You may find a particularly quiet moment in the day, either at work or at home. If you can only devote short time slots during the day and need several sessions to go through a paper, keeping a routine of daily reading is still advisable; find this time!

## Rule 2: Read thoroughly to build a sound background understanding of your topic

Particularly at early career stages, you need to get as much background information as possible. If you think you are short of time during your graduate period, just imagine how busy you will be later when you teach, supervise students, manage research projects; do not take shortcuts! Reading articles thoroughly will provide you a broad context about concepts, methods, results, and potential meanings and implications in your discipline. Once you have achieved a good background knowledge of a topic, you can start reading more selectively those papers or

Check for updates

OPEN ACCESS

**Citation:** Méndez M (2018) Ten simple rules for developing good reading habits during graduate school and beyond. PLoS Comput Biol 14(10): e1006467. https://doi.org/10.1371/journal.pcbi.1006467

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** October 11, 2018



**Funding:** The author received no specific funding for this article.

**Competing interests:** The author has declared that no competing interests exist.

paper sections that fill a particular gap or curiosity. But a good background knowledge takes a long time to build; do your homework early in your career so that you can save time later.

## Rule 3: Do not ignore the pillars of your discipline; read the classics

I repeat, read the classics! You may be working at the cutting edge of science, but your discipline surely has a long history behind it. It is naïve to think that you can go on in science by reading only what is new, starting at the moment you got involved in science, and ignore past fundamental steps in the building of your discipline (even if obsolete for current standards!). You need to identify and read the foundational papers of your discipline, concerning concepts, methods, or views. If this brings you back two centuries, so be it! Only by reading the classics you will get a deep understanding of your research topic. This will not only prevent you from reinventing the wheel but will enable you to build a robust context for significant advances. Repeat, read the classics!

## Rule 4: If you have to get familiar with a new topic, consider reading in chronological order

Early in your career, or whenever you get into a new topic and gather a bunch of references to get a hint on the status of the field, consider reading them in chronological order. If all references are recent, this gives you an idea of how the conversation is developing and who is answering who. If the time lapse is longer, you will get surprised to discover how concepts, methods, or interpretations may have changed with time. Reading in chronological order will allow you to realize subtle—or dramatic—changes in term use or meaning that will profitably add to your background knowledge of the topic.

## Rule 5: Avoid narrow-mindedness by reading beyond your discipline

Interesting ideas, concepts, methods, or implications are waiting for you in the work of other disciplines. Keep an eye to advances in other fields. You do not need to become the next Leonardo da Vinci; find a balance between focused and broad reading. The trick is to identify a few journals with a broad scope or that provide reviews in a broad spectrum of topics within a larger field and browse their content regularly (see Rule 6). You may also join journal clubs or follow renowned scientists outside your specific field using social networks.

## Rule 6: Create a list of relevant journals

Find out which journals publish relevant information for your research and subscribe to the online alerts for new content. Think big; if you find yourself checking less than 20 journals, you are probably missing relevant sources. Include journals that only occasionally publish relevant information. Do not forget those journals in which new analytical, experimental, or statistical methods are published, as they not always overlap with the journals in which primary research is published. Be prepared to reassess your list of journals during your scientific career; journals bust and fade and editorial lines change.

## Rule 7: Not all interesting stuff will appear in articles; read books

Books distil accumulated knowledge or suggest groundbreaking ideas. Read books, both classic and recent. Spotting the classic books is relatively easy and, with luck, you may be allowed to "plunder" the library of the elders around (Rule 10). However, finding new relevant books is not as straightforward as finding new articles. Google Scholar and most of the big editorials offer alerts or mailing lists for new books that can be customized. Another possibility is to identify the journals that include regular sections on new books, such as *Quarterly Review of Biology*. Finally, check with the library of your institution how new books are advertised—probably as electronic alerts—as the traditional shelf with novelties is almost extinct.

## Rule 8: Use a reference manager to keep track of your literature

Reference managers [6, 7] will avoid chaotic accumulation of nonretrievable literature. Some discipline is required to avoid misuse of reference managers. First, devote some time every day to update your recent readings. Second, never store in your reference manager papers you have not read; scientific databases already do that for you. Unread papers are best filed in a "to read" folder, that can be subdivided according to topics or urgency. Third, remember that storing a paper in a reference manager forces you to identify meaningful key words. Thus, refrain from merely importing key words already included in the reference and create your personalized list of keywords because this will contribute to building your background knowledge of the discipline.

## Rule 9: Keep a long-term review for your own use as a way to remember what you read

The more you read, the more you forget! A fruitful way to remember what you read is to open one (or several) review(s) of a topic for your own use. You may use your reference manager to do that (Rule 8). Nevertheless, I recommend going a step further and have a spreadsheet or a text document to store basic information or main messages from the primary literature. Whatever the format you choose, this review will allow you to retrieve much more easily confirmatory or negative evidence that you might use when writing an introduction or a discussion. At the same time, it will serve as a "note to yourself" for future surveys in the topic. Of course, it can eventually lead to a formal review paper.

## Rule 10: Build your own library to make yourself independent and inspire others

As soon as you can, start your own collection of papers and books. Do not rely exclusively on your supervisor's (or your colleagues' or your institution's) library. Certainly, the internet is there to store the information for you . . .except when it is not! The same applies to the library of your institution, particularly outside the top scientific countries where research funding is haphazard [8, 9]. Consider subscribing to some particularly interesting journal or becoming a member of a society that publishes a relevant journal. Building your own library in paper might be challenging early in your career due to funding limitations or movement between institutions during your postdoc stage. Nevertheless, applying this rule should not wait until you get a permanent position, and a digital library is always an option. For scientists at later stages in their career, a personal library should be considered a must, as an important part of mentoring. I might be a little old-fashioned, but for those of us who grew intellectually surrounded by books and journals in our departmental library—and plundered others' libraries—the mere presence of books around creates an inspiring atmosphere. Provide the same atmosphere for your fellow students or colleagues.

## Final remarks

Good reading habits are essential at all stages of a scientific career. Senior scientists should cultivate them, caress literature, and instill these habits to their students as a part of their mentoring. In particular, Rule 10 can have an important part in the mentoring process as a way of giving example and as a tool for sharing knowledge and curiosity. A good deal of self-discipline underlies good reading habits, especially rules 1, 2, 3, 5, 8, and 9. As self-discipline is an important skill for early career scientists, they should not forget to include reading as a way to develop this skill. Finally, science is about the excitement of discovery and the amusement of enhanced understanding. Reading contributes to both; reading is, my colleagues, fun!

## Acknowledgments

I thank Agus Ventre for very useful suggestions to improve this article.

## References

1. Deitering A-M, Ede L. Doing research: joining the scholarly conversation. In: Ede L, editor. The academic writer: a brief guide. New York: Bedford St. Martin's; 2011. pp. 155–216.
2. Bourne PE. Ten simple rules for getting published. PLoS Comput Biol 2005; 1(5): e057. https://doi.org/10.1371/journal.pcbi.0010057 PMID: 16261197
3. Rougier NP, Droettboom M, Bourne PE. Ten simple rules for better figures. PLoS Comput Biol 2014; 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833 PMID: 25210732
4. Mensh B, Kording K. Ten simple rules for structuring papers. PLoS Comput Biol 2017; 13(9): e1005619. https://doi.org/10.1371/journal.pcbi.1005619 PMID: 28957311
5. Bauer DC. Ten simple rules for searching and organizing the scientific literature. Nature Precedings, 2009; http://precedings.nature.com/documents/3867/version/1. https://doi.org/10.1038/npre.2009.3867.1
6. Larsen PO, von Ins, M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. Scientometrics 2010; 84: 575–603. https://doi.org/10.1007/s11192-010-0202-z PMID: 20700371
7. Hull D, Pettifer SR, Kell DB. Defrosting the digital library: bibliographic tools for the next generation web. PLoS Comput Biol 2008; 4(10): e1000204. https://doi.org/10.1371/journal.pcbi.1000204 PMID: 18974831
8. Salgado J, Alegre-Cebollada J, Daura X, Giráldez T. Spanish science funding: low and inefficient. Biofísica Magazine 2017; 7: 1–5.
9. Else, H. Latin American science funding crisis fuels brain drain. Times Higher Education, 2017, June 29 Available from: https://www.timeshighereducation.com/news/latin-american-science-funding-crisis-fuels-brain-drain

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for Lifelong Learning, According to Hamming

**Thomas C. Erren[1]*, Tracy E. Slanger[1], J. Valérie Groß[1], Philip E. Bourne[2], Paul Cullen[3]**

**1** Institute and Policlinic for Occupational Medicine, Environmental Medicine and Prevention Research, University Hospital of Cologne, Cologne, Germany, **2** Office of the Director, The National Institutes of Health, Bethesda, Maryland, United States of America, **3** Medizinisches Labor Münster, Münster, Germany

Philip E. Bourne is Founding Editor-in-Chief of *PLOS Computational Biology*
* tim.erren@uni-koeln.de

> "A mathematician, like a painter or poet, is a maker of patterns. If his patterns are more permanent than theirs, it is because they are made with ideas."—G. H. Hardy, "A Mathematician's Apology" [1]

Learning is a lifelong imperative for any scientist, and Richard Hamming provided timeless advice on how to achieve this. In this sequel to our 2007 contribution to the Ten Simple Rules series [2], we attempt to distil the essence of what this mathematician and computer science and telecommunications pioneer addressed in one of his talks [3] and in his book *The Art of Doing Science and Engineering*: *Learning to Learn* [4]. Hamming developed both the talk and the book as a synthesis of his graduate course in engineering at the United States Naval Postgraduate School in Monterey, California. We have organized his authoritative advice into ten rules. We believe these will equip the reader to more confidently face the unremitting emergence of an exponentially increasing amount of new knowledge, coupled with the equally relentless obsolescence of established knowledge, in a world containing a greater number of scientists than ever before. Our rules promote a certain "style of thinking." They also emphasize orientation towards the future and—we hope—will help the reader learn how to learn while motivating him or her to continue learning throughout life.



**Citation:** Erren TC, Slanger TE, Groß JV, Bourne PE, Cullen P (2015) Ten Simple Rules for Lifelong Learning, According to Hamming. PLoS Comput Biol 11(2): e1004020. doi:10.1371/journal.pcbi.1004020

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** February 26, 2015



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** The authors have declared that no competing interests exist.

## Rule 1. Cultivate Lifelong Learning as a "Style of Thinking" That Concentrates on Fundamental Principles Rather Than on Facts

As Hamming indicates, learning to learn depends on a certain style of thinking [4]. An important distinction here is between education and training: education is learning what you should do and when and why to do it, whereas training is learning how to do it. Obviously, to succeed you need to be both educated and trained. But in this process, it is important not to be a slave to facts. In 1968, David Horrobin, who interacted with Karl Popper, Linus Pauling, and Sir John Eccles, wrote that "the so-called facts which the student learns are at least five years out of date when he meets them and after ten years of medical practice they will be almost dead" [5]. As John Hunter has said, "too many facts crowd the memory without advantage" [6]. Therefore, when faced with new knowledge, try to establish nodal points in the knowledge network; these will help you to reconstruct the essence of what you have learned even if you can't remember the details. As Hamming said, "concentrate on fundamentals, at least what you think at the time are fundamentals, and develop the ability to learn new fields of knowledge when they arise so you will not be left behind" [4]. As you learn, watch out for "principles [6], "fundamentals" [4], and "patterns" (see the gem by Hardy [1] in the preface to this Editorial). These are likely to have much longer half-lives than the details involved. A further elaboration of this rule appears in [5]: "To paraphrase John Hunter, facts are important only in so far as they lead to principles: it is the principles which are vital and it is on these which you must concentrate. Fortunately, they are usually easier to remember than the facts themselves and also, once remembered, they allow the facts to be deduced from them". In other words, it is important to understand what your knowledge network is made for rather than knowing what it is made of.

## Rule 2. Structure Your Learning to Ride the Information Tsunami Rather Than Drown in It

Exponential growth of the amount of knowledge is a central feature of the modern era. As Hamming points out, since the time of Isaac Newton (1642/3-1726/7), the total amount of knowledge (including but not limited to technical fields) has doubled about every 17 years. At the same time, the half-life of technical knowledge has been estimated to be about 15 years [4]. If the total amount of knowledge available today is x, then in 15 years the total amount of knowledge can be expected to be nearly 2x, while the amount of knowledge that has become obsolete will be about 0.5x. This means that the total amount of knowledge thought to be valid has increased from x to nearly 1.5x. Taken together, this means that if your daughter or son was born when you were 34 years old, the amount of knowledge she or he will be faced with on entering university at age 17 will be more than twice the amount you faced when you started college.

To put it differently, during an average working life, the amount of valid knowledge in any field can be expected to at least double, while more than three-quarters of what we think we know now will have become obsolete. This process may well continue indefinitely [4]. In view of the additional fact that the human brain can process data no faster than about 60 bits per second [7], it is clear that we need some kind of structure to our learning if we are to ride the information tsunami [8] rather than drown in it.

## Rule 3. Be Prepared to Compete and Interact with a Greater and More Rapidly Increasing Number of Scientists Than at Any Time in the Past

As Hamming points out, despite a levelling-off in some of the richer countries [9], the number of scientists worldwide continues to increase rapidly, and it has been estimated that about 90% of the scientists who ever lived are alive today. There is no reason to expect that this trend is about to change any time soon [10].

## Rule 4. Focus on the Future but Don't Ignore the Past

As usual, Hamming put it trenchantly: "Teachers should prepare the student for the student's future, not for the teacher's past" [4]. Some teachers object to this, saying that no one can know the future. This is of course true but misses the point. We do not need to know what will happen—we just need to be ready when it does.

However, when riding the information tsunami into the future, try not to ignore the past entirely, as it may contain important information you may not be aware of. For example, an article in Nature under the headline "Where have I seen that before?" draws attention to the 103 years that elapsed between an experiment already done in 1908 and its accidental replication as the discovery of "new chemistry" in 206–2007 [11, 12].

## Rule 5. Look for the Personal Angle

Hamming uses anecdotes from his own life to highlight concepts about how to learn. Most of us do not move in exalted circles such as the research division of Bell Laboratories in Hamming's time, but this need not prevent us from looking for the personal angle, private detail, or scientific feud that may help us to anchor the key concepts in our minds within a meaningful and significant context.

In that vein, we may infuse our learning with personal details, which some scientists publish. Vivid exchanges between Watson, Crick, and Chargaff during and after the double helix was discovered are an example of this. Watson documented that Chargaff was irritated when Crick found it hard to "remember the chemical differences among the four bases" [13] as a guide to which one-to-one ratios suggested that things were likely to go together. In 1976, Chargaff recalled meeting the Nobel laureates Watson and Crick in 1952:

> . . .what I really was, when I first met the fervid pair in Cambridge, was baffled, for here were two people trying to fit the nucleotides into a helix and worrying about its pitch—it became a double helix, I believe, after I told them about our results—without bothering to look up the structures of the compounds they wanted to fit together [14].

Clearly, that Chargaff referred to his competitors inter alia as "two pitchmen in search of a helix" [14] (Pitchmen is an American term referring to hawkers such as those seen on TV shopping channels. These are always in search of some new "amazing discovery" to foist on a gullible public. Chargaff is also making a pun on the pitch of the double helix.) and as "scientific clowns" [13] provides a flavor of what can happen at the personal level in the "House of Science" [14]. Equally clearly, personal information can contribute to remembering details of the science involved.

## Rule 6. Learn from the Successes of Others

As Hamming says, because "there are so many ways of being wrong and so few of being right, studying successes is more efficient, and furthermore, when your turn comes you will know how to succeed rather than how to fail." In addition, he notes that "vicarious learning from the experiences of others saves making errors yourself" [4].

To exemplify, there is much we can learn from the successes of Sir John Eccles. He was wrong in expecting that synaptic transmission was electrical rather than primarily chemical. However, Eccles was right when, inspired by Popper's rigor to formulate testable hypotheses, he instigated experiments that ultimately contributed to falsifying the hypothesis of electrical synaptic transmission. Moreover, in 1963, Eccles was awarded a Nobel prize, together with Huxley and Hodgkin, for elucidating "the language of electrical nerve impulses and the responses of nerve cells engaged in replying to it at synapses" [15]. Taken together, Eccles could be viewed as an example that rigorous scientific reasoning and work qualify for the "few [ways] of being right" [4], irrespective of the hypothesis you investigate. Making this your work mode may yield the strongest recognitions for scientists, i.e., even the Nobel prize.

## Rule 7. Use Trial and Error to Find the Style of Learning That Suits You

Hamming says that learning how to learn is like learning how to paint. Both need advice but also individual trial and error:

> How to be a great painter cannot be taught in words; one learns by trying many different approaches that seem to surround the subject. Art teachers usually let the advanced student

> paint, and then make suggestions on how they would have done it, or what might also be tried, more or less as the points arise in the student's head—which is where the learning is supposed to occur! [4]

Your teachers are thus important, but do not be afraid to think for yourself. To cite Horrobin again, "perhaps the most important thing for a student to realize is that while his teachers may differ from him in experience, with rare exceptions they will not differ in intelligence. They are not individuals to fear, but individuals to question and challenge" [5].

## Rule 8. No Matter How Much Advice You Get and How Much Talent You Possess, It Is Still You Who Must Do the Learning and Put in the Time

To quote Hamming, "I am. . .only a coach. I cannot run the mile for you" [4]. You must test how Hamming's advice in these ten rules works for you. You will only get out of these suggestions as much as you put in. Grapple with the issues involved and compare what is offered here with your own experience and that of others. Select and adapt those points that help you.

A prerequisite, of course, is native talent. But even for the talented, no amount of utilizing smart methods can substitute for sheer duration of effort. Gladwell has suggested that about 10,000 hours of application are needed to become a true expert in a particular field [16]. While some have quibbled with the universal validity of this suggestion, we think it is a fairly good estimate of what you need to put in. Also, these long hours need to be quality time without distractions. Easy to say, hard to do.

## Rule 9. Have a Vision to Give You a General Direction

A key to learning how to learn is to be economical and to structure your efforts according to the general direction in which you want to move. Hamming writes: "It is well known the drunken sailor who staggers to the left or right with $n$ independent random steps will, on the average, end up about $\sqrt{n}$ steps from the origin. But if there is a pretty girl in one direction, then his steps will tend to go in that direction and he will go a distance proportional to $n$." [4]

You too need an attractive vision of where you want to go. As Hamming points out, "having a vision is what tends to separate the leaders from the followers."

## Rule 10. Make Your Life Count: Struggle for Excellence

Overall, Hamming is "preaching the message that, with apparently only one life to live on this earth, you ought to try to make significant contributions to humanity rather than just get along through life comfortably—that the life of trying to achieve excellence in some area is in itself a worthy goal for your life.. . .[A] life without a struggle. . .is hardly a life worth living."

The true gain is in the struggle for excellence, and "a life without such a goal is not really living but it is merely existing" [4].

## Acknowledgments

TCE thanks Robert Yates for reminding him that Hamming offered much insight into how to research and learn. The authors wish to thank Ingrid Vukadinovic and Peter Hulajko for their dedicated library support.

## References

1. Hardy GH (1940) A Mathematician's Apology. Cambridge: Cambridge University Press. PMID: 20783471

2. Erren TC, Cullen P, Erren M, Bourne PE (2007) Ten simple rules for doing your best research, according to Hamming. PLoS Comput Biol 3: 1839–1840. PMID: 17967054
3. Hamming RW (1995) Intro to The Art of Doing Science and Engineering: Learning to Learn. http://www.youtube.com/watch?v=AD4b-52jtos&list=PL2FF649D0C4407B30. Accessed 28 November 2014.
4. Hamming RW (1997) The Art of Doing Science and Engineering—Learning to Learn. Amsterdam: Gordon and Breach Science Publishers. PMID: 25165803
5. Horrobin DF (1968) Medical Physiology and Biochemistry. London: Edward Arnold Ltd. PMID: 4309973
6. Hunter J (1835) The works of John Hunter, FRS. Palmer JF, editor. Vol. 1. London: Longman et al.
7. MIT Technology Review (2009 August 25) New measure of human brain processing speed. Emerging Technology from the arXiv. http://www.technologyreview.com/view/415041/new-measure-of-human-brain-processing-speed/. Accessed 28 November 2014.
8. Erren TC, Cullen P, Erren M (2009) How to surf today's information tsunami: on the craft of effective reading. Med Hypotheses 73: 278–279. doi: 10.1016/j.mehy.2009.05.002 PMID: 19493630
9. Cyranoski D, Gilbert N, Ledford H, Nayar A, Yahia M (2011) Education: The PhD factory. Nature 472:276–279. doi: 10.1038/472276a PMID: 21512548
10. UNESCO (2010) UNESCO Science Report 2010. The current status of science around the world. http://unesdoc.unesco.org/images/0018/001899/189958E.pdf. Accessed 28 November 2014. PMID: 25506974
11. Sanderson K (2007 December 4) Where have I seen that before? 103-Year-old chemical reaction pops up again. Nature News. http://www.nature.com/news/2007/071204/full/news.2007.341.html. Accessed 28 November 2014.
12. Christl M (2007) 1,7-Diaza[12]annulene derivatives? 100-year-old pyridinium salts! Angew Chem Int Ed Engl 46: 9152–9153. PMID: 18046689
13. Watson JD (1968) The Double Helix. A Personal Account of the Discovery of the Structure of DNA. New York: Atheneum.
14. Chargaff E (1974) Building the tower of Babble. Nature 248: 776–779. PMID: 4599083
15. Granit R (1963) Award Ceremony Speech. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1963/press.html. Accessed 28 November 2014.
16. Gladwell M (2008) Outliers: the Story of Success. New York (New York): Little, Brown and Company. PMID: 25506952

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Online Learning

**David B. Searls***

Independent Consultant, Philadelphia, Pennsylvania, United States of America

The success of online courseware such as that offered by the Massachusetts Institute of Technology (MIT) (http://ocw.mit.edu) and now by many other institutions, together with a plethora of recent announcements of major new initiatives in this arena such as Coursera (https://www.coursera.org), Udacity (http://www.udacity.com), and the Harvard-MIT partnership edX (http://www.edxonline.org), have made it clear that online learning has reached a tipping point. Many signs point to the possibility in the near future of getting a quality, university-level education at a distance, and for free. As exciting as this prospect may be, it behooves online students to follow a few simple rules for getting the most out of the experience, while being realistic in their expectations, as outlined below.

## Rule 1: Make a Plan

There are many possible motivations for becoming involved in online learning, whether in bioinformatics or any other field. There's nothing wrong with taking an online course on impulse, or to fill a very specific need, or simply for fun, if that's your goal. But if you hope to acquire a broader swath of knowledge for some larger purpose, you will need a directed, organized approach to be efficient and effective, especially in the absence of a formal degree program or traditional academic advisor. Don't underestimate the importance, or the difficulty, of this planning effort, particularly if you are new to the field.

Start by deciding on a curriculum that suits your needs, and determining the optimum sequence of courses. The companion article "An Online Bioinformatics Curriculum" by this author offers a selection of courses and some advice on possible tracks within that particular interdisciplinary field [1]. More generally, professional societies will often publish curriculum recommendations, and yet another approach is to examine university course catalogs (often available online) for their recommended curricula in your area of interest. Do this for several programs that you know to be of high quality, to arrive at a consensus. The particular courses needn't be online—you just need the titles or descriptions so that you can seek out an analogous course that is available to you. Pay attention to the syllabus and listed prerequisites to make sure the online course is at the appropriate level, and that you have adequate preparation. If no online version is available, the same search will allow you to determine an appropriate textbook and/or reading list to make up the gap.

Finally, keep in mind what you hope to achieve at the end. This may be simply knowledge for its own sake, but if you need a set of skills to accomplish some life goal (and in particular if you need it to help you get a job, gain advancement, or qualify for some academic program), that fact should shape not only which courses you take, but how you take them. This will be discussed further in Rule 9.

## Rule 2: Be Selective

We generally try to get best value-for-money in our education, as in all else. As the number of courses available continues to grow, you may have the opportunity to shop around. Be selective in choosing online courses, because in fact there's no such thing as a free course, to the extent that you value your time. Seek out the best institutions, and then seek out the best teachers, just as you would on campus. If you don't already know a professor by reputation, it's easy to do the usual assessment by means of bibliographic search. A web search of a particular course identifier at a major university will typically yield course/professor evaluations by students in various fora, though in practice these have a tendency to generate more heat than light.

The evaluation of a course must also take format into account. In many cases only bare, unannotated videos of lectures are available, as for the UC Berkeley Webcast resource (http://webcast.berkeley.edu), and these can be highly variable in quality and especially in legibility of projected slides or writing on a blackboard or whiteboard. A quick sampling will help you decide if the video quality suits you. If not, try searching for a corresponding course website to see if the slides or detailed course notes are separately available, so that you can have them up while you run the video (or even audio, which in rare instances is the only form of recording available).

In other cases, no video is provided but only detailed syllabi, lecture notes, readings, quizzes, exams, and/or demonstrations (as for the majority of courses on the MIT OpenCourseWare (OCW) project (http://ocw.mit.edu), and innumerable course web sites). To be sure, these materials can be valuable, and in particular there is a growing trend to posting notes in highly polished form and even as full-fledged online textbooks. But video lectures have many advantages: a sense of immediacy, the feeling of a personal touch, helpful emphasis and nuance in the presentation, and the simple fact that a memorable professor makes for memorable subject matter. In such skilled hands, the video format affords the use of techniques that have been shown to enhance learning, including not only visual material but also expressions of enthusiasm by the lecturer and even humor [2].

Having both lecture videos and ancillary course material, as for example with the Open Yale courses to varying degrees

**Citation:** Searls DB (2012) Ten Simple Rules for Online Learning. PLoS Comput Biol 8(9): e1002631. doi:10.1371/journal.pcbi.1002631

**Editor:** Fran Lewitter, Whitehead Institute, United States of America

**Published** September 13, 2012



**Funding:** The author received no specific funding for this article.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: David.B.Searls@gmail.com

David B. Searls is an Associate Editor of *PLOS Computational Biology*.

(http://oyc.yale.edu), is a huge plus. Even better is the recent trend to structured courseware in which video presentations are modularized and interspersed with assessments and homework. Some of these are even run on a set schedule, so that videos are released and homeworks are due on a regular weekly basis. All else being equal, the courses offered under this model by initiatives such as Coursera and edX should always be preferred.

## Rule 3: Organize Your Learning Environment

Create an environment that will promote learning. In terms of time, a regular schedule (such as those imposed by the structured initiatives) can help ensure that you keep momentum, so that successive learnings reinforce each other before they have a chance to fade. Let your mind get used to a regular workout. And resist the temptation to rush through an online course; cognitive psychologists studying learning have long noted a "spacing effect" that suggests it is better to absorb material at regular, separated intervals than all at once, which is why "cramming" for exams is so ineffective [3].

In terms of your work space, use common sense and give yourself some quiet and privacy to concentrate on lectures. Set up your screen with an enlarged video display together with associated materials such as slides, and close other windows, especially e-mail and the like. Cognitive psychologists now have experimental evidence of the insidious effect of multitasking on the learning process [4]. While online courses often provide detailed lecture notes, it is a good idea to take your own notes anyway, as research has long suggested the benefits of this activity to long-term learning, particularly at deeper levels of understanding [5].

## Rule 4: Do the Readings

There is a whole set of truisms about classroom education that carries over to online education with very little change, except perhaps the need for a greater emphasis when one is on one's own. Particularly important is to prepare for each class as indicated in the syllabus, which generally means completing the required readings ahead of time. It's too easy to neglect this when online, knowing you will never be called on in class by the instructor.

Don't adopt the attitude that the lecture should be a painless way to be spoon-fed the material and that otherwise you may as well just read the textbook on your own. In almost every case the lecturer is doing you the favor of assigning carefully selected readings rather than the entire book, helping you to focus on the "meat" that is most relevant to the approach being taken in the course. Your end of the bargain is to come prepared so that you can most efficiently absorb the value-added of the multimedia presentation.

Many courses include journal articles in their assigned reading, as a teaching tool. Particularly as you reach more advanced stages, it's a good idea to read the current literature on your own as well, not only for the learning experience but because it will reveal to you any gaps in your knowledge or skills. This in turn may prompt you to adjust your selection of courses or independent study. At the end of your efforts, the ability to read and understand the key journals in your field will establish both your competency and currency in the subject.

## Rule 5: Do the Exercises

Another platitude that will come as no surprise is that you should do your assigned homework before you get to watch any TV. You know in your heart that you learn by doing, and once again it's too easy to neglect this when nobody is collecting and grading your work.

For courses that are computational in nature, there is the additional imperative to do the programming assignments faithfully. You haven't really taken a programming course if you haven't been through the hard slogging: designing, testing, debugging, documenting, refactoring, etc. If the assignment is just too dull, you have the luxury of being able to tweak it towards some variation that interests you, but come what may, it's well known that you must put in the time to become a proficient coder.

In fact, if you are doing general programming classes in an interdisciplinary program like bioinformatics, you may want to put in a special effort to modify the programming assignments and projects so that they apply to your domain of interest. You will need experience obtaining and working with the relevant data as much as you need practice with the programming languages and algorithms themselves. It's hard enough being interdisciplinary without doing twice the work, so kill two birds with one stone whenever you can.

## Rule 6: Do the Assessments

Many courses will have quizzes and exams in addition to homework, and once again it is important not to neglect these, for the obvious reason that you need to know how well you are absorbing the material. This is the opportunity to make mid-course corrections, which are far easier to execute in online learning. Moreover, a well-constructed exam can be a learning experience in itself, particularly if it is the occasion for you to whip up your competitive instincts just a bit, since this is likely to improve your retention. Many lines of research point to the benefits of testing in promoting effective learning [6].

The newer structured learning approaches such as Coursera not only institutionalize exams, running them in set intervals, but also integrate quizzes directly into lectures. This format promises to provide feedback not only to the users but to the providers as well, indicating what points need clarification in the associated discussion fora or supplemental short videos. No doubt there will be other ways in which data analytics can be applied to improving the educational experience and effectiveness of these new platforms.

## Rule 7: Exploit the Advantages

There is at least some experimental evidence that online lectures with integrated assessment can produce results superior to the classroom for large courses, possibly by avoiding known issues of "poor attendance… and inappropriate behavior (talking, sleeping, reading) on the part of students who do attend class" [7]. A meta-analysis of ten such comparative studies found that online versions of courses had learning outcomes better than traditional versions in four cases, worse in two, and indistinguishable in four [8]; the authors of this study noted pros and cons to online learning: "The significant pro is the element of convenience which eliminates the constrictive boundaries of space and time. The most notable con involves the impersonal nature of the online environment."

The disadvantages of online learning will be addressed below. As for the advantages of convenience, first among these is the flexibility you have to schedule classes, in whatever order is appropriate. Most courses today are available to download and view as you see fit, and at a pace that suits your schedule. Gone are the days of missing out on an upper-level elective because it is only offered in alternate years, or the professor is on sabbatical. If you start a course and find you are ill-prepared, you can "drop" the course without prejudice and look for the appropriate prerequisite. While it is true that the newer structured courses are often offered on a set schedule, indications are that the materials will still be available afterwards for individual viewing, though without some of the ''live'' features.

Even the viewing experience itself offers new possibilities. One can freeze a lecture to work out a knotty point, and back up to repeat sections. Entire lectures can be repeated as needed, for example after any sort of forced hiatus. Many courses recorded in classrooms spend a major portion of the first lecture and parts of others dealing with administrivia, which is easily skipped over. Many video players used in these sites allow for a lecture to be sped up by 1.25× or 1.5×, for lecturers who speak glacially or for segments that are easily grasped. In short, learning online gives the student unprecedented control over the proceedings.

## Rule 8: Reach Out

It is widely believed that the greatest disadvantage of online learning relates to its isolated and impersonal nature [9]. This perspective is ironic given that the Internet is also known as the ultimate medium for social networking. In fact, the newer structured approaches like Coursera do have discussion forums that allow students to commune with one other, which in addition are monitored by instructors who can post authoritative responses to questions and even alter elements of the course as needed. Take advantage of these features instead of just lurking. If you are taking a course without such a built-in community, it will take more work, but if you can locate or recruit other like-minded students you may find that you learn better in a group, however small.

A more difficult challenge is the absence of a live instructor or tutor with whom you can directly interact. Having someone to coach you through a difficult concept, targeting your specific needs, is an advantage of traditional education that is obviously lacking in recorded lectures. Research is underway in adaptive instructional systems [10] that may eventually contribute to the new online learning initiatives, but in the meantime you must make the best of it.

For some, the most valuable feature of live instruction may not be the continual give-and-take so much as the ability to get past some particular roadblock to understanding. In such cases one can always try contacting the course instructor with a very specific question, though obviously you should not be surprised if this tactic fails. You can also try cold-mailing other teachers, researchers, or experts, preferably with some connection to you. In the end, though, you are most likely going to puzzle it out for yourself, through extra reading and online search. Don't be too discouraged by this, because the time and effort you put in will make it more likely that you retain the concept and understand it at a deeper level; educational psychologists call this ''productive failure'' [11].

## Rule 9: Document Your Achievements

A somewhat cynical view of traditional higher learning is that its value lies in the diploma as much as the knowledge. Certainly the lack of a diploma from any self-study program, online or otherwise, can be an impediment to getting the opportunity to apply one's hard-won knowledge. To a potential employer or a graduate admissions office, for example, a diploma and a transcript are proof of sorts that you know what you say you know.

In its current state, open online learning simply doesn't have an equivalent mechanism. Recent programs have taken to offering certificates of completion, but these are carefully distinguished from actual course credits at the sponsoring institution. There are several movements to establish formal systems of so-called badges indicating skills and achievements acquired from activities like online learning, particularly in computational circles, for instance in the case of Mozilla Open Badges (http://openbadges.org). Any of these may or may not carry some weight in the outside world, depending on a given employer's or university's trust in both the effectiveness and the integrity of the online resource, but they certainly don't approach the luster of a degree from an accredited school.

Whether or not any sort of certification is connected with courses you take, if you want to demonstrate that you have learned the material you should take concrete steps from the outset to document this in any way you can. Record your program of study in something resembling a transcript, including brief course descriptions, institutions, and dates of completion. Create a portfolio of accomplishments, including reading lists, major exams, and any substantial homework assignments; while it may not be possible to prove definitively that you came by them honestly, it will suggest that you were serious about your studies. Perhaps add some work on your own initiative, such as essays or projects based on learnings from the course. For computational courses in particular, assemble your programming assignments and projects, making sure that the code is impeccably documented and presentable. Even if such a portfolio of accomplishments lacks the imprimatur of a formal diploma, it is a good habit of mind and can be motivational as well.

## Rule 10: Be Realistic

It is important to be realistic in your expectations of online learning. As Rules 3–6 suggest, the single most important factor in your success at learning will be your degree of motivation, which in turn will determine your receptiveness and work ethic. While motivation is also necessary for success in a traditional campus environment, there is no doubt that learning on your own requires a special brand of persistence, and is not for everyone. Don't set yourself up for disappointment by mistaking free courseware for a free lunch. On the other hand, a casual attitude is fine as long as it aligns with your expectations.

Be cognizant, also, of the inherent limits of online learning. As discussed in Rule 8, without live instructors and tutors you will encounter various kinds of hurdles that, however, may be overcome with extra effort. What is harder to surmount is the lack of an academic advisor to guide you on the broader issues. Rule 1 offers general advice on curriculum planning, but true mentorship, extending to career advice, is harder to come by. Jump at every chance you get to ask experts for guidance, however fragmentary, and pay attention to the literature to be sure you are studying what is really called for.

Other limits to online education include the lack of certain kinds of training, such as laboratory experience and presentation skills. These shortcomings may be a bit less critical in some areas, such as computer science, but they will loom larger in fields like molecular biology. In such cases you will need to use your imagination to fill the gap; for instance, you might seek unpaid internships or consider paying for some carefully selected traditional classes.

In the final analysis the problem of certification raised in Rule 9 may be the single biggest practical issue facing the online learner. John F. Kennedy, upon being awarded an honorary degree from Yale, said that ''now… I have the best of both worlds, a Harvard education and a Yale degree'' [12]. One is left to wonder how he would have fared with a Harvard education and neither degree. If you are happy to have the knowledge without the diploma, then you may rest easy on this point, but otherwise you need to do a careful cost-benefit analysis before investing time in an online education.

While it is important to be realistic about all these factors, don't let it dampen your enthusiasm. With the steadily increasing momentum behind online learning, it is more than likely that the initiatives themselves will begin to address the issue of effective certification, and perhaps even that of mentorship, even as the catalog of available courses expands rapidly. Anyone beginning a course of online study today is likely to find a greatly changed landscape even by the time they complete their education.

## References


1. Searls DB (2012) An online bioinformatics curriculum. PLoS Comp Biol 8: e1002632. doi:10.1371/journal.pcbi.1002632.
2. DeWinstanley PA, Bjork RA (2002) Successful lecturing: presenting information in ways that engage effective processing. New Directions for Teaching and Learning 89: 19–31.
3. Bahrick HP, Bahrick LE, Bahrick AS, Bahrick PE (1993) Maintenance of foreign language vocabulary and the spacing effect. Psychological Science 4: 316–322.
4. Foerde K, Knowlton BJ, Poldrack RA (2006) Modulation of competing memory systems by distraction. Proc Natl Acad Sci U S A 103(31): 11778–11783.
5. Bohay M, Blakely DP, Tamplin AK, Radvansky GA (2011) Note taking, review, memory, and comprehension. Amer J Psychol 124(1): 63–73.
6. Roediger HL, Putnam AL, Smith MA (2011) Ten benefits of testing and their applications to educational practice. In: Mestre J, Ross BH, editors. Cognition in education, volume 55 (Psychology of Learning & Motivation). Burlington, MA: Academic Press. pp. 1–36. Available: http://psych.wustl.edu/memory/Roddy%20article%20PDF%27s/BC_Roediger%20et%20al%20%282011%29_PLM.pdf. Accessed 13 August 2012.
7. Maki RH, Maki WS, Patterson M, Whittaker PD (2000) Evaluation of a Web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. Behavior Research Methods, Instruments, and Computers 32(2): 230–239.
8. Weber JM, Lennon R (2007) Multi-course comparison of traditional versus Web-based course delivery systems. J Educators Online 4(2). Available: http://www.thejeo.com/Volume4Number2/Weber%20Final.pdf. Accessed 26 July 2012.
9. Brooks D (2012) The campus tsunami. The New York Times. 4 May 2012. p. A29. Available: http://www.nytimes.com/2012/05/04/opinion/brooks-the-campus-tsunami.html. Accessed 26 July 2012.
10. Moller L, Harvey DM (2009) Learning and instructional technologies for the 21st century. New York: Springer. 248 p.
11. Kapur M, Bielaczyc K (2012) Designing for productive failure. J Learning Sci 21(1): 45–83.
12. Kennedy JF (1962) Commencement address at Yale University, New Haven, Connecticut, 11 June 1962. John F. Kennedy Presidential Library, accession number WH-104. Available: http://www.jfklibrary.org/Asset-Viewer/Archives/JFKWHA-104.aspx. Accessed 26 July 2012.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for providing effective bioinformatics research support

**Judit Kumuthini[1]*, Michael Chimenti[2], Sven Nahnsen[3], Alexander Peltzer[3], Rebone Meraba[1], Ross McFadyen[1], Gordon Wells[1], Deanne Taylor[4], Mark Maienschein-Cline[5], Jian-Liang Li[6], Jyothi Thimmapuram[7], Radha Murthy-Karuturi[8], Lyndon Zass[1]**

**1** H3ABioNet, Centre for Proteomic and Genomic Research, Cape Town, South Africa, **2** Iowa Institute of Human Genetics, Bioinformatics Division, Carver College of Medicine, University of Iowa, Iowa City, United States of America, **3** Quantitative Biology Centre, Eberhard Karls University of Tübingen, Tübingen, Baden-Württemberg, Germany, **4** Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **5** Research Informatics Core, University of Illinois at Chicago, Chicago, Illinois, United States of America, **6** Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences, Durham, North Carolina, United States of America, **7** Bioinformatics Core, Purdue University, West Lafayette, Indiana, United States of America, **8** Department of Computational Sciences, The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, United States of America

* jkumuthini@gmail.com



OPEN ACCESS

**Citation:** Kumuthini J, Chimenti M, Nahnsen S, Peltzer A, Meraba R, McFadyen R, et al. (2020) Ten simple rules for providing effective bioinformatics research support. PLoS Comput Biol 16(3): e1007531. https://doi.org/10.1371/journal.pcbi.1007531

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** March 26, 2020



**Data Availability Statement:** There is no associated data with the manuscript.

**Funding:** The authors are supported by various grants. This include the H3ABioNet grant, supported by the National Institutes of Health Common Fund under grant number U41HG006941. S. Nahnsen also acknowledges funding from the Deutsch Forschungsgemeinschaft (DFG) grants: SFB/TR 209 "Liver cancer" and Project ID 398967434 – TRR 261. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Abstract

Life scientists are increasingly turning to high-throughput sequencing technologies in their research programs, owing to the enormous potential of these methods. In a parallel manner, the number of core facilities that provide bioinformatics support are also increasing. Notably, the generation of complex large datasets has necessitated the development of bioinformatics support core facilities that aid laboratory scientists with cost-effective and efficient data management, analysis, and interpretation. In this article, we address the challenges—related to communication, good laboratory practice, and data handling—that may be encountered in core support facilities when providing bioinformatics support, drawing on our own experiences working as support bioinformaticians on multidisciplinary research projects. Most importantly, the article proposes a list of guidelines that outline how these challenges can be preemptively avoided and effectively managed to increase the value of outputs to the end user, covering the entire research project lifecycle, including experimental design, data analysis, and management (*i.e.*, sharing and storage). In addition, we highlight the importance of clear and transparent communication, comprehensive preparation, appropriate handling of samples and data using monitoring systems, and the employment of appropriate tools and standard operating procedures to provide effective bioinformatics support.

## Author summary

The article we wrote draws from our experience in core support facilities and highlights 10 best practices that individuals who apply information technology approaches to

**Competing interests:** The authors have declared that no competing interests exist.

biological, medical, and health research should consider when providing support to individuals who generate data for this research in the lab. As interdisciplinary approaches are increasingly being utilized within the biological and medical sciences, effective collaboration and support between the aforementioned parties is crucial to promote the quality and integrity of research. These practices highlight the importance of quality control, comprehensive reporting, effective communication, and more in the production of quality data as well as the promotion of effective collaboration.

## Introduction

Because of the technological boom, life scientists are increasingly turning to high-throughput sequencing in their research programs and generating enormous volumes of data [1]. These projects are characterized by the use of specialized computational and tools to analyze the generated data, highlighting the need for interdisciplinary services and/or deep collaborations between primary data-generating researchers and bioinformaticians [1]. This trend has resulted in the establishment of both commercial and departmental (core) bioinformatics support facilities worldwide [2]. Because these facilities provide support to data-generating researchers in their data analysis and reporting, bioinformaticians in these facilities may inevitably encounter erroneous datasets (i.e., low-quality datasets primarily caused by experimental failures such as inadequate experimental design, improper sample collection and processing, sample contamination, degradation, sequencing, hybridization, library preparation, equipment and reagent failures, and more). When faced with erroneous data, bioinformaticians may be left without the necessary resources to address the associated challenges (*e.g.*, which analysis method to employ). In essence, this highlights the importance of effective collaboration between bioinformaticians and data-generating researchers to provide effective support and analysis [3].

In this addition to the "Ten Simple Rules" series, we propose 10 rules to facilitate bioinformaticians in providing effective research support. These rules were developed based on extensive experiences of bioinformaticians working in core facilities and ordered to reflect the natural sequence of events in a project's lifetime (project development, data collection and generation, and data analysis).

These rules can be scaled to both small single-site and large collaborative research projects and are therefore discussed as such. With the understanding that core facilities receive research projects at different stages of the project lifecycle, not all rules can always be implemented; however, these rules represent best practices that should be followed as much as possible to ensure the quality and integrity of all data collected and generated within a given research project. By implementing the following rules, bioinformaticians who routinely provide bioinformatics support to data-generating researchers can work to establish more realistic expectations for analysis while improving the quality and value of outcomes, owing to improved communication, experimental design, record keeping, data management and analysis. In addition, these rules discuss how to prevent the production of erroneous data as well as how such data can be treated.

## Project development

### Rule 1: Collaboratively design experiment

Successful bioinformatics analyses are dependent on appropriate experimental design, as previously described [4]. A good experimental design starts with a well-defined hypothesis and covers sample strategies (e.g., number and frequency), data handling, and data reporting. The experimental design should aim to reduce the types and sources of variability, increase the generalizability of the experiment, and make it replicable and reusable [4]. It is both easier and more cost efficient to identify and correct experimental design issues ahead of time than to address deficiencies thereafter. Thus, discussion between data-generating researchers and bioinformaticians is highly desirable and should occur as early as possible during project development and experimental design. Even so, bioinformaticians may not always have the luxury to provide input on experimental design. In such cases, it may be beneficial to request the experimental design and highlight concerns that may be of significance during data analysis.

During the experimental design discussions, a number of issues should be addressed, including cost, confounding batch effects, effect size, technical and biological replicates, sample integrity and purity, and controls. Researchers may be tempted to conduct many comparisons within the framework of one experiment containing large sample sizes (typically by sacrificing biological replicates). Therefore, it is crucial to discuss the critical role of appropriate sample sizes and replicates (biological and technical) [5], gain an understanding of the variables being investigated, and discuss the importance of avoiding confounding batch effects. If multiple samples or conditions are included in a project, batches should be constructed in a manner that evenly or randomly distributes experimental conditions across all the batches and processes during each experimental stage [6]. Similarly, the expected effect size of the test conditions should be carefully considered, as researchers may make unspoken assumptions about model system alterations while failing to plan for adequate replication to measure small effects [7].

## Rule 2: Manage scope and expectations

Successfully executed experiments are associated with attentive experimental design and clear communication [8]. Like Rule 1, communications regarding the potential limitations and pitfalls of a project (including technology, resources, and analysis) should occur prior to conducting the experiments. These communications should strive to eliminate extraneous technical detail without oversimplifying the topics (providing appropriate reference materials where required) [8]. Topics that should be covered include the employed wet and dry laboratory workflows (transparency should be provided from both sides) and, to avoid dissatisfaction, the expected and realistic turnaround times (it may be beneficial to clarify that these estimates refer to the time following receipt of data). In these initial communications, it is crucial to clarify the methods and responsible persons of future communications. To ensure that communications are clear and effective, a written analytical study plan (ASP) outlining the aforementioned topics should be prepared and agreed upon by all involved parties. Employed workflows should be documented and appropriately shared to enable bidirectional knowledge transfer for future reference. The ASP should be comprehensive and refer to the experimental design. It should also include the agreed upon timelines, the exact deliverables, and an alternative plan, in case the original data analysis plan is deemed insufficient.

To provide effective support and deliver the scientific vision of a project, scope management is critical [9]. The primary scope management patterns to monitor are (1) "scope grope," in which a project takes an undefined path with no sight of completion, resulting in wasted resources without impact; (2) "scope swell," in which the project expands rapidly without thoughtful allocation of resources and time, resulting in stress on the core and affecting the number of other projects which can be supported; and (3) "scope creep," in which a project expands slowly but significantly, resulting in delayed project delivery, loss of impact, and over-consumption of planned resources. To flexibly manage the scope of a project and the expected outcome, universal adoption of the project management methodologies (including organizing resources, setting key milestones, and communicating to-go/not-to-go plans) is crucial and one of the primary aims of the developed ASPs. Ultimately, all involved parties should understand the proposed research vision and associated methodologies. The ASP serves to promote (1) easy sharing and storing of the study information and experimental design and (2) easy tracking of the project from wet to dry laboratory. Clear communication is thus imperative to providing effective support because it enables mutual knowledge transfer and understanding.

### Rule 3: Define and ensure data management

When receiving new support projects, it is critical to define the scope of data management required and set measures to ensure this management. A comprehensive data management plan (DMP) can be used to achieve this in projects involving high-throughput technology and data generation. Typically, a core facility should have a general standard operating procedure for data management that covers data handling during active analysis followed by long-term storage and backup. Because a core support facility typically deals with data owned by another party, a core's responsibility may only extend to secure storage of data and results for the client (short and long term), whereas final data sharing is the data owner's responsibility. In many cases, however, data-generating scientists may call upon bioinformaticians to facilitate the development of a DMP for a research endeavor. As bioinformaticians in core facilities, it is crucial to communicate the importance of comprehensive DMPs to data owners. Similarly, bioinformaticians should be aware and communicate the extent of their core's DMP.

Guidelines to developing a good DMP have been previously described [10]. Aspects to consider when developing a DMP include determining the legal, ethical, and funder's requirements associated with the data; identifying the types of data to be collected; identifying the standards and ontologies that will be employed; and determining how data will be organized, quality controlled, documented, stored, and disseminated [10]. In addition, core facilities should also consider defining applicable data handling policies and preparing data management budgets.

Comprehensive DMPs aim to address the ethical, governance, and resource requirements associated with the data; promote findable, accessible, interoperable, and reusable (FAIR) research [11]; and consider associated data security, access, and the backup concerned. Ultimately, the DMP provides assurance for the long-term preservation and accessibility of the generated data [12]. Like the experimental designs, DMPs can be collaboratively developed or selected by data-generating researchers and bioinformaticians. However, bioinformatics core facilities may also choose to develop a standard DMP that can be adjusted as required for individual projects.

## Data collection and generation

### Rule 4: Manage the traceability of data

Traceability of all samples and data in a research project is a crucial component of effective bioinformatics support [13]. Traceability should be comprehensive and encompass sample acquisition and processing, as well as data generation, analysis, storage, and reporting [13]. The best-case scenario is a database management system that is maintained by and accessible to both the data-generating scientists and bioinformaticians. In practice, this kind of system is called a laboratory information management system (LIMS) and may be implemented to simplify the traceability of samples and data, thereby reducing human error and the production of erroneous data [14]. If a LIMS is not feasible, a shared cloud-based resource may serve the same purpose [15].

These systems should enable the production of reliable results at a faster rate than manual systems and enable data tracking from sequencing runs over time and across experiments in order to improve efficiency and trace down potential errors [14]. Additionally, these systems also promote quality control by highlighting failed samples and identifying the accountable parties. Enabling sample and data traceability is ultimately one of the most efficient ways to identify sources and prevent production of erroneous data [14]. Notably, bioinformaticians may not always be part of a sequencing core and are therefore dependent on data owners providing accurate information. Comprehensive DMPs (see **Rule 3**) may need to account for the precise setup applicable to individual clients.

## Rule 5: Determine how and what metadata are reported

In order for bioinformaticians to conduct appropriate downstream data analysis of an experiment, the associated metadata must be provided. Metadata should be as complete as possible and should include the experimental variables of interest, all aspects of sample handling, known or suspected sources of batch variables, and laboratory mistakes such as sample mislabels and swaps [16]. To account for the aforementioned considerations, an effective system should be implemented to ensure comprehensive metadata reporting. Ideally, data-generating facilities should adopt a system to enable tracking of critical information related to the experiments and pass this information to the bioinformatics core (see Rule 4). Because many research groups may not have the luxury of an LIMS, the data-generating researchers and bioinformaticians should propose or develop standardized worksheets or web-based submission forms for metadata reporting, which designate required and optional fields [16]. In the absence of a standardized approach, metadata reporting may be provided in various forms (e.g., spreadsheets, handwritten notes, etc.); however, these often lack critical batch information and other insights from the wet laboratory experimental practices. If possible, it might be worth planning from the onset where data will be made publicly available. In many cases, the data will be deposited in an existing public repository; therefore, knowing the structure and depth of metadata collection required for the repository is crucial. For smaller-scale studies, metadata templates provided by the repository can be used to record samples so that everything is already prepared for final submission as well.

In the interest of producing interoperable research, metadata reporting should adhere to experiment-specific reporting guidelines, such as Minimum Information About a Microarray Experiment (MIAME) [17], Minimum Information required for a DMET Experiment (MIDE) [18], Minimum Information About a Proteomics Experiment (MIAPE) [19], and more. These can be accessed through FAIRsharing (https://fairsharing.org/) (a standards-housing resource), BioSchemas (https://bioschemas.org/), and the Global Alliance for Global Health (GA4GH) (https://www.ga4gh.org/). Similarly, metadata ontologies can be found in online repositories such as Bioportal (https://bioportal.bioontology.org/) and Ontology Lookup Service (https://www.ebi.ac.uk/ols/index). Selecting and employing appropriate reporting standards should be covered in the DMP (**Rule 3**) and may be required by journals or funders. Reporting standards ensure that researchers adhere to internationally set standards during their experimental procedures. Moreover, employing data reporting standards, helps to promote reuse and comparison to previously conducted studies [20]. Ultimately, this ensures that both researchers and their community reap the maximum benefit from their collected and generated data.

### Rule 6: Coordinate data and internet security

Providing assurances that data are both secure and stable is an important aspect of providing effective bioinformatics support [21]. Although these aspects are typically addressed by an information technology (IT) department or system administrator, it is crucial to communicate the particular requirements with the responsible person(s) and may be an important consideration in resource- or capacity-limited facilities.

Data security refers to the prevention of harmful cyber-attacks and unoptimized internet security issues, as well as the setting of data access and transfer limitations [21]. Generally, the individuals with access to research data should be limited to parties with relevant responsibility and accountability. Cases pertaining to personal data, particularly patient data, may require auditing of data access as well. The aspects that need to be considered when safeguarding data to maintain quality, include (1) confidentiality (maintaining access and transfer); (2) integrity (ensuring information is accurate, valid, and reliable); (3) availability (resources and support are available); (4) accountability (actions can be attributed to relevant parties); and (5) provenance (origin and history of data are known and well defined).

Internet security refers to the use and stability of the internet, which is employed to manage and analyze data associated with high-throughput experiments [21]. To address the computational challenges (e.g., central processing units [CPUs], memory, storage) associated with high-throughput data analysis, cloud computing has emerged as the leading solution. In these cases, the importance of ensuring data and internet security are further emphasized. As a result, cloud users have to rely heavily upon the service providers for data privacy and security protection; therefore, data backups and recovery plans should be maintained and monitored.

GA4GH has released a data security toolkit (www.ga4gh.org/genomic-data-toolkit/data-security-toolkit/) for genomics and health-related data sharing. This toolkit consists of recommendations for privacy and security safeguards and procedures for maintaining proper access and fidelity of data. Useful tips to support this maintenance include (1) developing access control documents (that are reviewed and updated periodically); (2) implementing data verification and reporting processes; (3) implementing risk management strategies; (4) establishing strong working relationships with local IT support; (5) implementing regular maintenance and upgrade processes; and (6) implementing real-time server monitoring systems and maintaining security certificates associated with maintained sites and software [21].

## Data analysis

### Rule 7: Control data quality throughout the project lifecycle

Quality control is inarguably the most important component of high-throughput experiments. This pertains to both the quality control of data generated by high-throughput technologies to enable downstream analysis as well as the quality control of the generated results to make reliable scientific inferences. Quality control occurs throughout the project lifestyle (i.e., during each implemented standard operating procedure of a workflow). In some cases, experimental failures may be inevitable; therefore, data quality control needs to be performed by the bioinformatics core at various stages of processing. In commercial facilities, quality control is typically ensured by a quality control manager, who maintains the quality control processes, conducts root cause analyses, and implements corrective and preventive actions [22,23]. In core facilities, quality control is the responsibility of all bioinformaticians.

Documenting the implemented quality control procedures is a crucial component of this rule [3]. When reviewing data quality, it is essential for a bioinformatician to be able to refer to the quality control procedures implemented to appropriately interpret the metrics and, subsequently, conduct suitable analysis. Importantly, quality control needs to be implemented at both sample level and cohort level. Whereas the former identifies inadequate sample data, the latter identifies outliers in the overall data of the cohort. Following the interpretation, bioinformaticians should effectively communicate quality metrics to primary investigators, to identify potential issues, make go or no-go decisions, and design the proper analytical approaches for addressing their research objectives. The selection of appropriate quality control processes, gates, and values play an important part in the downstream analysis of high-throughput omics data [24]. Where necessary, it may be useful to follow developers' recommendations. Such processes eliminate or reduce erroneous data within a data set and may be adjusted to salvage as much data as possible. Balancing the data quality parameters and statistical power is key, thus, one should proceed with caution. With this in mind, the implementation of Rules 1 and 2 is crucial to appropriately handle these cases, especially in high-throughput research, as the selection of quality control processes and parameters needs to be appropriately justified in research communications. The incorporation of the previous rules aims to facilitate quality control, which highlights the importance of implementing this rule to maintain research and experimental integrity.

## Rule 8: Identify suitable computational tools for data analysis

An important component of providing effective bioinformatics support is conducting research that is reproducible and reusable. When conducting data analysis, it is crucial to employ appropriate bioinformatics methods (tools and resources) and statistical models that deliver reliable inferences from the data. As is the nature of the science, several bioinformatics tools have been developed and proposed for application in high-throughput experiments. However, no tool is expected to be the best for all situations, though tools can be recommended for repeated or common workflows. So how do we determine an appropriate tool to use? Firstly, we have to size up the data characteristics with the aims of the analysis; attention needs to be given to the strengths and limitations of a given tool for the analysis at hand. Analysis using tools that are of academic standard are usually a good place to start; however, we can also look to which tools are employed by similar projects. In addition, several other features may be investigated to identify appropriate tools; these include whether the tool is supported by the developers, whether the tool gains active support in relevant question and answer (Q&A) forums, whether the tool is open source, documented, and version controlled, and, depending on the bioinformaticians' experience, whether the tool is easily installable, executable, and parallelizable. It may be useful for a core facility to have a procedure or criteria in place for the use of new tools when analyzing high-throughput data.

Notably, when implementing a selected method, significant attention needs to be given to the measurement of $p$-values and estimating false discovery rates (FDRs) due to the violation of assumptions of statistical models and dependency among the hypotheses tested [25, 26]. Ultimately, the use of specific tools, statistical models, and values adds an important layer of understanding to the overall research project. This enables and promotes future collaborations, allows others to critically evaluate the research at hand, and increases the credibility of the findings and allows the researchers themselves to identify the limitations and strengths of their research and generated data.

## Rule 9: Track, record, and confirm workflow changes

Establishing methods to track and record changes to workflows can go a long way in improving bioinformatics support services and ensuring quality control during data analysis. In practice, this is called verifying and validating workflow changes and is typically required to adhere to international quality management standards, such as those proposed by the International Organization for Standardization (www.iso.org/home.html).

Because of technical or software updates, adjusted project requirements, or process improvements, workflows may be altered from time to time. Whenever such alterations occur or new workflows for specific analyses are developed, it is important to independently verify and validate them. Technical verification and validation are not only necessary to ensure that new or altered workflows are working as expected and are fit for purpose but also to ensure that the workflow can be maintained while handling data inputs of different sizes and types and adapting to different technical landscapes [27]. Validating workflow alterations and communicating these alterations to collaborators or clients are essential for reproducible research and scope management, as described in Rule 2.

Deviations refer to any observed events in data analysis procedures that are exceptions or alterations from specifications or acceptance criteria [28]. These may include out of specification, tolerance or trend results, deviations from an approved standard operating procedure, test method, validation protocol or ASP, and software failures [28]. Maintaining a system by which these deviations can be reported and monitored functions as an important component of both metadata reporting and quality control and maintenance [23]. These reports supplement root cause analysis and corrective and preventive action in commercial facilities, as described in **Rule 8**. These practices may also prevent consistent production of erroneous data and simplify error tracking. A bug tracking and change management system would be critical in core facilities in which multiple people may be working on complex workflows/pipelines at the same time. Ultimately, such practices also provide assurances in the core facilities' practices and capacity to collaborators and clients.

## Rule 10: Repurpose the data

In cases wherein erroneous data are produced, researchers may choose to terminate a project to save research funds or conform to service agreements. However, the data produced may yet be informative. Before terminating a project, there should be clear communication (as outlined in Rule 2) between the bioinformaticians and primary researchers; the cost of the experiments may be weighed up against the outputs that may still be desirable and relevant to the end user, highlighting the importance of effectively communicating the pros and cons of the decision. A detailed sample, design, and tool review may inform the aforementioned decision. A sample review includes the review of LIMS data, the cohort, the batches, the adaptors, and identifiers. This review aims to identify where experimental failures occurred or where erroneous data were produced. On the other hand, a design review includes cohort composition analysis, power analysis, and batch identification and confounding. This review aims to identify faults within the experimental design; these may be adjusted or tightly regulated in the future. Lastly, in a computational tool review, tools are verified and validated using test data, and maintenance and suitable support for the tools are identified.

Importantly, with regards to low-quality data or marginal data (datasets closer to the lower limit of qualification and acceptability, i.e., datasets that barely exceed the minimum requirements for downstream analysis), there are 2 ways in which projects can be continued: (1) using partial data or (2) repurposing the data. The former refers to the use of replicates with enough depth and quality to answer the initially posed research questions, albeit at a smaller scale. In case of the latter, although the data generated may not be sufficient for answering the initial research question, it may be appropriate to repurpose the data by answering an additional or alternative research question within the scope of the project. Importantly, marginal data can also be used for improvement of workflows, procedures, and overall quality of similar studies in the future and could be used to guide future experimental procedures and designs.

## Conclusion

High-throughput data play a key role in expediting scientific discoveries and rapidly providing scientific understanding to improve human health [1]. Therefore, effective collaborations with bioinformatics cores are essential to modern bioscientific research. Theoretically, data-generating and bioinformatics cores may be seen as 2 separate entities, functioning separately in the same research project; however, they are intrinsically connected and highly dependent on each other to function effectively. Effective bioinformatics collaborations aim to conduct quality research and reduce the production of marginal data. To meet these aims, these collaborations require clear communications between the 2 entities of the collaboration; appropriate reporting and documentation that can be referred to in the future; the appropriate collection and reporting of data and metadata; appropriate quality control, validation, verification, and deviation reporting procedures; and the use of appropriate technology and computational tools that are specific to both the data generated and the research questions being investigated. Although majority of the rules apply to maintaining and ensuring data quality, taking the same approach to data exploration and analysis stages can result in analyses that are inflexible and might miss important but unexpected findings. Having an analysis structure that is resistant to change can tend to prefer a stock analysis rather than adapting to the early stage findings. Building in consultation checkpoints between bioinformaticians and data-generating scientists through all stages of the project lifecycle is crucial to ensure that the best results are obtained. Again, the best rule to adopt and implement will depend on the nature of the study, but pointing out that some parts of the analysis are easier to predefine than others might be a useful addition.

Overall, effectively implementing any of these rules in bioinformatics support facilities will facilitate increased productivity, credibility, and satisfaction while simultaneously reducing erroneous data production and promoting high-quality research. Ultimately, the proposed rules ensure that information is reported and communicated correctly, at the highest quality, making it broadly beneficial.

## References


1. Luo J, Wu M, Gopukumar D, Zhao Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. Biomed Inform Insights. 2016 Jan 19; 8:1–10.
2. Lewitter F, Rebhan M. Establishing a Successful Bioinformatics Core Facility Team. PLoS Comput Biol. 2009 Jun; 5(6):e1000368. https://doi.org/10.1371/journal.pcbi.1000368 PMID: 19557124
3. Casadevall A, Ellis LM, Davies EW, McFall-Ngai M, Fang FC. A Framework for Improving the Quality of Research in the Biological Sciences. mBio. 2016 Sep 7; 7(4):e01256–16. https://doi.org/10.1128/mBio.01256-16 PMID: 27578756
4. Curtis MJ, Bond RA, Spina D, Ahluwalia A, Alexander SPA, Giembycz MA, et al. Experimental design and analysis and their reporting: new guidance for publication in BJP. Br J Pharmacol. 2015 Jul; 172 (14):3461–3471. https://doi.org/10.1111/bph.12856 PMID: 26114403
5. Yang Y, Fear J, Hu J, Haecker I, Zhou L, Renne R, et al. Leveraging biological replicates to improve analysis in ChIP-seq experiments. Comput Struct Biotechnol J. 2014; 9:e201401002. https://doi.org/10.5936/csbj.201401002 PMID: 24688750
6. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010 Oct; 11(10):733–739. https://doi.org/10.1038/nrg2825 PMID: 20838408
7. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. J Grad Med Educ. 2012 Sep; 4(3):279–282. https://doi.org/10.4300/JGME-D-12-00156.1 PMID: 23997866
8. Wong-Parodi G, Strauss BH. Team science for science communication. Proc Natl Acad Sci U S A. 2014 Sep 16; 111(Suppl 4):13658–13663.
9. Mirza MN, Pourzolfaghar Z, Shahnazari M. Significance of Scope in Project Success. Procedia Technol. 2013 Jan 1; 9:722–729.

10. Michener WK. Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol. 2015 Oct 22; 11(10):e1004525.
11. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15; 3:160018. https://doi.org/10.1038/sdata.2016.18 PMID: 26978244
12. Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS. Data management in clinical research: An overview. Indian J Pharmacol. 2012 Mar; 44(2):168–172. https://doi.org/10.4103/0253-7613.93842 PMID: 22529469
13. Csavina J, Roberti JA, Taylor JR, Loescher HW. Traceable measurements and calibration: a primer on uncertainty analysis. Ecosphere. 2017 Feb 1; 8(2):e01683.
14. Cucoranu IC. Laboratory Information Systems Management and Operations. Clin Lab Med. 2016 Mar; 36(1):51–56. https://doi.org/10.1016/j.cll.2015.09.006 PMID: 26851664
15. Kyobe S, Musinguzi H, Lwanga N, Kezimbira D, Kigozi E, Katabazi FA, et al. Selecting a Laboratory Information Management System for Biorepositories in Low- and Middle-Income Countries: The H3Africa Experience and Lessons Learned. Biopreservation Biobanking. 2017 Apr 1; 15(2):111–115.
16. Hong EL, Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, et al. Principles of metadata organization at the ENCODE data coordination center. Database (Oxford). 2016 Mar 15;2016: baw001. https://doi.org/10.1093/database/baw001 PMID: 26980513
17. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet. 2001 Dec; 29(4):365–371. https://doi.org/10.1038/ng1201-365 PMID: 11726920
18. Kumuthini J, Mbiyavanga M, Chimusa ER, Pathak J, Somervuo P, Van Schaik RH, et al. Minimum information required for a DMET experiment reporting. Pharmacogenomics. 2016 Sep; 17(14):1533–1545. https://doi.org/10.2217/pgs-2016-0015 PMID: 27548815
19. Taylor CF, Paton NW, Lilley KS, Binz P-A, Julian RK Jr, Jones AR, et al. The minimum information about a proteomics experiment (MIAPE). Nat Biotechnol. 2007 Aug; 25(8):887–893. https://doi.org/10.1038/nbt1329 PMID: 17687369
20. Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. PLoS Biol. 2018 Nov 20; 16(11): e2006930. https://doi.org/10.1371/journal.pbio.2006930 PMID: 30457984
21. Kantarcioglu M, Ferrari E. Research Challenges at the Intersection of Big Data, Security and Privacy. Front Big Data. 2019; 2. https://doi.org/10.3389/fdata.2019.00001
22. Manghani K. Quality assurance: Importance of systems and standard operating procedures. Perspect Clin Res. 2011; 2(1):34–37. https://doi.org/10.4103/2229-3485.76288 PMID: 21584180
23. Badrick T. The Quality Control System. Clin Biochem Rev. 2008 Aug; 29(Suppl 1):S67–70.
24. Leggett RM, Ramirez-Gonzalez RH, Clavijo BJ, Waite D, Davey RP. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. Front Genet. 2013 Dec 17; 4:288. https://doi.org/10.3389/fgene.2013.00288 PMID: 24381581
25. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016; 31:337–350. https://doi.org/10.1007/s10654-016-0149-3 PMID: 27209009
26. Chen JJ, Roberson PK, Schell MJ. The false discovery rate: a key concept in large-scale genetic studies. Cancer Control. 2010 Jan; 17(1):58–62. https://doi.org/10.1177/107327481001700108 PMID: 20010520
27. Antonelli G, Padoan A, Aita A, Sciacovelli L, Plebani M. Verification or validation, that is the question. J Lab Precis Med. 2017 Aug 13; 2(8). https://doi.org/10.21037/jlpm.2017.07.11
28. Ghooi RB, Bhosale N, Wadhwani R, Divate P, Divate U. Assessment and classification of protocol deviations. Perspect Clin Res. 2016; 7(3):132–136. https://doi.org/10.4103/2229-3485.184817 PMID: 27453830

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for providing optimal administrative support to research teams

**Romina Garrido** [1,2‡*], **Casandra A. Trowbridge** [3‡], **Nana Tamura** [4]

**1** Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Catalonia, Spain, **2** Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain, **3** Department of Genetics, Stanford School of Medicine, Stanford University, Stanford, California, United States of America, **4** RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan

‡ These authors share first authorship on this work.
* romina.garrido@crg.es



**Citation:** Garrido R, Trowbridge CA, Tamura N (2019) Ten simple rules for providing optimal administrative support to research teams. PLoS Comput Biol 15(10): e1007292. https://doi.org/10.1371/journal.pcbi.1007292

**Editor:** Fran Lewitter, Whitehead Institute for Biomedical Research, UNITED STATES

**Published:** October 3, 2019



**Funding:** We want to thank Roderic Guigó for providing funding from the Universitat Pompeu Fabra (UPF) for the visit to the Riken and to the Spanish Ministerio de Educación, Cultura y Deporte, programa "Salvador de Madariaga" PRX17/0039. CRG acknowledges the support of the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership and the Centro de Excelencia Severo Ochoa, and to the CERCA Programme / Generalitat de Catalunya. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Conducting science nowadays is not only properly devoting time to research but also managing administrative tasks. With increasing amounts of time needing to be invested into these administrative tasks, it leads to less time devoted to science and research itself. According to a *Nature* survey, research group leaders point out having more institutional support for administrative tasks as one of the most important needs they have and that institutions need to provide this support to lab members in addition to the leaders [1,2]. Proper administrative support can have a remarkable impact on the scientific productivity of research groups by contributing to a more efficient and effective work environment. It allows the scientists to focus mainly on the science.

With our experience as administrative-based staff in international research groups located in Japan, Europe, and the United States, we will lay out a number of rules that can provide guidance on where to apply efforts when assisting research teams in academic and research environments. These efforts are typically different from what is experienced when supporting teams in business and industry atmospheres. In research, people are used to freedom of thought and to sometimes challenging authority and questioning rules or any preestablished concepts. You need to be much more patient and flexible, and you don't get desperate if things are different from what you were used to if you are coming from a different environment. These rules are mainly addressed to administrators directly supporting research teams, such as secretaries, assistants, project managers, research managers, operations managers, or any other similar job position. These roles typically have direct and immediate impact on the daily activity of research groups. These rules may also help the principal investigators and other research leaders understand the contributions and positive impact administrators have within a research group.

## Rule 1: Boost your "soft skills"

You will need a number of technical and specialty skills to perform your daily administrative activities, including "soft skills" [3,4]. Soft skills refer to the more intangible and nontechnical abilities. According to the Research Administration as a Profession (RAAAP) Worldwide project survey, examples of important soft skills are your professional capabilities tied to collaboration, taking responsibility, adaptability, problem solving, and communication [5]. Soft skills influence the manner in which you handle interaction, multitasking, and leadership with a team. These skills are more personality driven and can be a sign of emotional intelligence.

Soft skills come naturally for some people, but these skills can also be learned. Developing your soft skills requires a great deal of people interaction and practice. You can boost these skills by looking for mentors, participating in working groups to hear different perspectives, learning various approaches to tasks, and getting enrolled in training courses and workshops in and outside of your home institution. There are associations that offer educational and skill development resources aimed specifically for research administrators. Some examples include the Society of Research Administrators International (SRAI), ARMA International (formerly the Association of Records Managers and Administrators), European Association of Research Managers and Administrators (EARMA), and the National Council of University Research Administrators (NCURA).

People with strong soft skills are more likely to be successful in research and can contribute to a proficient multilateral research environment.

## Rule 2: Be proactive and be decisive

Adopt a proactive mindset. Tackle challenges and solve problems. Measure a desired outcome and take the initiative to see it through. Taking initiative demonstrates that you are a self-starter, independent, and highly motivated. Being proactive with your work is an opportunity to stand out and excel in your job while increasing the likelihood of stability, progress, and success for your research team and the science being conducted.

It is extremely useful to have the skill of being able to anticipate the needs of the research team, especially the principal investigator. Research teams are very typically busy with managing multiple projects and grants, writing new grant proposals, attending meetings, giving frequent presentations, facilitating lectures, and traveling quite often, among other tasks. Strong coordination efforts and keeping updated common calendars where all these tasks can be visible at a glance will help to keep competing commitments, priorities, and deadlines from overlapping. Helping them stay on time and ahead on tasks is greatly needed. Anticipating what they need ahead of time without being told is invaluable.

Be decisive in your actions. You do not want to overthink the most basic of decisions. Research team members and colleagues need to be able to rely on a person who is assertive and can make confident decisions than to put their trust in someone who is slowed and paralyzed by indecision. When a big decision seems like it could be too much to tackle all at once, break it down, take small steps, get more information, reconsider, and then make the next decision [6].

These characteristics described show your leadership and strategic thinking capabilities and that you are willing to go the extra mile in your work. These qualities are extremely useful to supporting your research team and ultimately help excel the progress of science.

## Rule 3: Be efficient, effective, and communicative

You play a vital role in keeping research teams organized in an efficient and effective manner. This includes helping your team remain mindful of timelines and deadlines, supporting the management of competing tasks, developing engagement strategies, scheduling all necessary meetings, keeping costs minimal, and defining success metrics. It also includes minimizing distractions for yourself as well as for members of the research team. The goal is to have optimal effectiveness and enhance functionality for the whole research team.

Your role greatly helps researchers navigate all the different aspects of administration, including managing requirements, obstacles, and burden. Examples include the following:

- Grants have requirements to fulfill. You can help your principal investigator on the oversight of grant rules and procedures, keeping protocols up to date, writing reports, and fulfilling requests from the Institutional Review Board (IRB).
- Research units with higher bureaucratization have lower scientific performance [8]. You can help identify priorities and action plans, eliminate paperwork and extra processes whenever you can, and avoid pushing out decisions or bottlenecking progress.
- Scientists are sometimes reluctant to accept new administrative procedures and/or to spend funds on administrative activities [7]. You can help ease administrative changes and burden by streamlining and automating tasks whenever possible and independently handling what you can on your own. You can help your principal investigator manage administrative budgets and recommend where funding should be allocated.
- Metrics and indicators of team activities are important in research. Although keeping track of some of the items—such as the number of publications and citations using the Altmetric tool—is relatively easy, it may be more difficult for some others [8]. Important team activities to track include the following: current positions held and who are alumni (to evaluate the success on training) and actions taken for public engagement and communication of science (all forms of scientific communications, which include talks to general audiences, radio and TV interviews, articles in newspapers, blog posts, etc.).

Being efficient and effective cannot happen without communication, communication, and communication. Research administrative roles are typically tasked to manage multiple forms of communication. In addition to being an effective verbal communicator, excellent writing skills are necessary. It is essential to write reports, taking minutes of meetings as well as contacting people by email about a range of issues from funding to organizing a program of lectures [9]. There may be promotional leaflets, course materials, exhibition programs, and prospectuses to be written, proofread, and edited [9].

We encourage you to attend lab meetings, working group meetings, and consortium meetings. Although these meetings may be mainly focused on science and analysis, it allows you to hear updates on projects and gives you the opportunity to engage your research team and ask questions. You can be the point person for identifying best practices and communication plans for administrative tasks, increasing efficiency and effectiveness for the entire research team.

## Rule 4: Collaborate and network with other administrators

Although you are directly supporting researchers and your relationship with them is very important, you also need to nurture the relationship with administrative staff located in other departments of your institution. You will very likely find yourself working with the departments leading finance and grants administration, procurement, events and marketing, information technology, human resources, facilities management, etc. You will need to rely on these departments and their services to successfully streamline needs and solve issues for your research team. In many cases, you are the point person, or the "bridge," through which all communication funnels between researchers and other departments. Interacting with other members of administration at your institution will also help contribute to creating an efficient work environment, as described in Rule 3.

Not only is it important to build relationships within your institution, but it is also important to do so with external collaborators. This is certainly the case whether you work on a research team that spans across multiple institutions or is a part of a large consortium. These types of groups require large-scale coordination efforts, which administrative support professionals are typically tasked to handle. You may find yourself working with administrators from each institution involved in the same study or project as yours. If there are consortium meetings, consider attending them and have administration-based parallel meetings. You can ask the principal investigator to budget for these in grant applications. All these efforts will help you feel like you are part of the scientific enterprise.

Networking can also help your career path in science. This can help us think “outside of the box” and see a bigger picture. Visiting other institutions to learn about how they manage their administrative tasks can be beneficial. It could also lead to new collaborations and other professional opportunities. This “Ten Simple Rules” paper is a great example of the result from professional relationships through networking and collaborations spanning over the last 6 years.

Being able to establish strong collaborations, network, have positive rapport, share resources, and learn from other disciplines are valuable benefits in facilitating your job and coordinating the needs of your research team. This will likely open new doors professionally.

## Rule 5: Be curious about science, even if you do not have a scientific background

When directly supporting scientists working in a particular research field, it can be useful to have some background and exposure in that same scientific discipline to accompany your administrative experience. Although this is ideal for hiring administrators in scientific research fields, it is not easy to find job candidates with such a background, and thus it cannot always be a prerequisite for hiring. When administrators are just starting their career in academia for the first time, they must learn how to adapt to the workplace traditions, customs, and tolerances, especially if they are coming from a career in industry and traditional business in which working standards are very different. Regardless, what is important is for the administrator to be curious about science and make the effort to get a basic familiarity with the scientific discipline and academic working environment they are supporting.

The best way you can learn about the scientific field you are involved with is to have frequent interaction with the scientists you work for. Have scientists explain to you what they do in a nontechnical manner and don’t be afraid to ask questions. Many researchers are used to explaining what they do in layman terms from attending various meetings and giving presentations to audiences with wide-ranging skill sets and educational backgrounds.

We encourage you to get enrolled in basic science or fundamental research courses at a university, whether they are in-person or online courses. Check with your institution on available tuition reimbursement plans or financial assistance programs to enroll in these types of courses. If this type of institutional support does not exist, be proactive about asking your employer to organize these financial assistance programs.

You can also attend scientific meetings, seminars, and conferences aimed at general audiences. Participating in activities where you can interact with people of different professional and educational backgrounds will be very beneficial for you.

You should also take time to learn about your research team’s behaviors and tolerances, as well as your institution’s workplace customs. Academia tends to be more informal than traditional business atmospheres and offers scheduling flexibility, autonomy, freedom to collaborate, and cross-disciplinary thinking. On the flip side, researchers are under immense pressure to be self-starters, continually publish their research, promote and advocate for their work, and find funding sources. Take some time to find how your working style fits into the academic world and how to increase your chances of success.

As mentioned in Rule 1, there are professional associations that offer educational and skill development resources for research administrators—SRAI, ARMA, EARMA, and NCURA. These associations have done extensive work in laying the professional development framework (PDF) and organizing educational opportunities for research administrators. For example, ARMA ran a 12-month project to produce a well-researched and evidence-based PDF for research managers and administrators [10]. From these findings, ARMA developed an outline for professional qualifications and developed courses, certification programs, and online learning opportunities that can contribute to the maturity of the profession, which is very different depending on the regions of the world [11]. Although we focused on the work done specifically by ARMA, joining any of these professional associations will provide guidance on the research administrative career field and give you access to plenty of courses and webinars.

Spending time on getting a better understanding of what's done in the scientific field you're involved in will be a very wise and smart investment for you, your research team, and your institution. It also allows you to build a robust and long-term career in the research environment.

## Rule 6: Be responsible with data sharing and handling

Working with principal investigators, research teams, managers, stakeholders, and sponsors means you will be privy to a wealth of information, including highly sensitive and confidential information. This can include exposure to unpublished data and personal health information (PHI). Accessing, using, and/or distributing such sensitive information without permission could give rise to unwanted and serious consequences. Knowing what information to share and with whom is a must, and your discretion needs to be trusted. Make sure to have a clear understanding of when to share the data, with whom, in what format, what security measures are in place, and how the data transfer will be handled. Ultimately, principal investigators, research managers, and institutions are responsible for educating and training employees on the access and handling of sensitive information, as well as providing information on the ethical issues involved. There should be protocols and procedures in place. Be sure to know them and to enroll in any course that can provide you trainings on these matters.

If the study you support works with human subjects and PHI, your research team will be working with the IRB for approval to conduct the study. At times the IRB can feel like an oppressive oversight body bound by regulations and designed to inhibit research, but the IRB is in place to protect human subjects from unethical scientific research while ensuring the highest quality research standards [12]. You can administratively support and soften the relationship between the IRB, the protocol director/principal investigator, and the research team. Administrative roles typically can help with IRB correspondence, protocol modifications and amendment submissions, and preparing reports.

Try to be familiar with the rules and directives governing the sharing of information. Because the principal investigator in your group may not have time to familiarize with the directive, try to investigate how this can affect the science carried out in the group. It will also help your research team maintain compliance.

## Rule 7: Participate in the onboarding process as much as possible

The onboarding process for newcomers is a crucial time for getting them settled into their new position, with their new team, and within the organization as a whole. Although human resources, principal investigators, and lab managers are a large part of the onboarding process for a new employee, administrative staff are typically involved in the process as well. It is important for new employees to feel they are supported by their new organization, especially because starting a new job can cause some challenges. Challenges can be more significant for a newcomer who has taken a job in a foreign country, which is common in research.

Ahead of the new employee's arrival, connect with the principal investigator and lab manager(s), as well as human resources, to make sure that an onboarding plan is in place with delegation of who will be responsible for each step and task. Providing new employees with a streamlined onboarding process affects all aspects of success for a team and organization as a whole. Examples of action items you can do for new lab members could include the following: being a point person for any questions the new hire may have, assisting with getting the new hire signed up for any relevant onboarding courses or certifications, informing the group about the new hire so they are aware of his/her arrival, and looking for a mentor who can introduce the newcomer to the rest of the lab members and inform him/her about any relevant information. Although all these examples may seem like small tasks and gestures, they can go a long way in making your new hire feels warmly welcomed, valued, and set up for success.

The golden rule to keep in mind is to treat someone the way you would like to be treated, meaning treat a newcomer, at both the professional and personal level, the way you want to be treated if you were in the same situation. You will also have a good working relationship with the new lab member from the very beginning.

## Rule 8: Appreciate and support cultural diversity

There are multiple educational and work opportunities abroad in science and research. It is very common to have research teams with wide-ranging cultural diversity. Expectations for behavior in areas such as leadership, communication, and feedback style can vary across cultures [13]. It is important to be sensitive to cultural differences and to avoid inadvertently stereotyping [13]. Administrative staff, alongside the principal investigator and research management, can actively play a role in learning about the different cultures that exist within the research team and how to avoid misunderstandings and miscommunications. Administrative staff can help define clear expectations for administrative-based actions and tasks while assisting the incoming international team member with overall adaptation with the research team and new work environment.

Culturally related work preferences, such as time management, task orientation, risk orientation, directness, or even sense of humor, have an important impact on the team dynamics. Some internationally based scientists find that the priorities attached to socializing (including the interaction between men and women) differ from what they are used to [13]. They will very likely feel the difference in thoughts and values and often encounter challenges, particularly with the relationship to authority. These views may differ strongly between countries, and you could help the newcomers to adapt to the accepted rules in the institute.

To learn about your team's various cultural working and social preferences, you can begin by asking them where they are from and institutions they have studied and worked at. Ask about their communication preferences and the management working styles that they are used to when handling tasks with administration, project coordination and management, and overall team interactions and management. You can be the person they can turn to when difficulties arise at work, especially if they feel uncomfortable asking those questions with their lab mates or other colleagues. Encourage them to keep an open mind and ask questions. The goal is to work towards finding a common understanding and establishing clear expectations. In the process, you may end up learning a new method or way of doing something to apply to your own job or with the team as a whole.

You can also recommend ways for them to learn more about the local culture. You can tell them about local traditions and holiday festivities. You can suggest restaurants they should check out and new food they should consider tasting. You can suggest places where they can hear locally appreciated music or museums to visit. Immersion in the local culture is a great way for them to adapt to their new community.

It is important to value cultural diversity and understand differences between people while recognizing that these differences are a valued asset. Multiculturalism improves productivity, creativity, and employee engagement and opens up doors to new opportunities in often unexpected ways [14,15]. Blending and cocreating workflows and methods can help you and the whole team create a rich, balanced, and comfortable environment. Be engaged with the cultural diversity on your research team.

## Rule 9: Treat everyone fairly

On a research team, there are people at very different levels of hierarchy, responsibility, and needs. Administrative staff are typically assigned to give more priority to supporting the requests and needs of the more senior members (faculty, principal investigators, senior postdocs, etc.), but those in earlier career stages (undergraduates, technicians, junior postdocs, etc.) need support as well. Regardless of seniority and status, everyone on the research team needs to feel he/she is receiving adequate administrative support and being treated fairly. They do not all need the same level of support, but the quality of the services you provide to all team members should be similar.

Administrative support resources are not unlimited, so it is good to set a list of services that you can offer to people at each level of responsibility and to review it periodically to make any necessary changes. Do not limit yourself to that list, as you always want to be looking for new ways to expand and grow your skill set. Make sure that it is understood that more senior level requests and urgent matters will always take priority but that you will also address other requests. It is helpful to let people know expectations and timelines related to task completion. Periodically check in with each team member to see if they need assistance with any tasks. Your administrative support benefits the research team as a whole.

People are always the most important asset and should receive excellent service. Everyone should have the feeling that their contributions to the team and to science are important. Once they become alumni, they will be the ambassadors of the institution and of your work.

## Rule 10: Be an active team player and show your unique qualities

Although your contribution to the research team is not scientific, you play a vitally significant role to help the principal investigator streamline the working dynamics of the group. Don't stay in the shadows or be a mere observer. Work actively to find ways to make administrative procedures easier for your research team. Give your point of view and suggestions for improvements and help develop more effective methods. Engage your research team on their thoughts. Combining both administrative and scientific perspectives when facing problems will inspire you to find creative solutions to support your research team.

Organizations are more than their mission statement, aims, impact, infrastructure, and policies. Organizations are greatly defined by the type of people they hire, which creates the company culture. Show your unique qualities. Let your personality shine through. Be genuine. Everyone in research appreciates character, so be yourself.

Your active team player attitude and unique qualities will have great positive effect on your research team as well as on the institution as a whole. Take pride in that.

## Conclusion

Administrative professionals do more than just assist. They are one of the backbones to successful research and organizations. They have strong skills for multitasking, planning, organization, customer service, and directing large groups of people. They are effective communicators and trusted individuals to handle sensitive information. They are the gatekeepers to senior level individuals and the helpers to the whole team. They are the bridges to other departments and help foster collaborations with other researchers and labs. They take initiative and do what is needed without being told and can anticipate future needs. Their roles in science are constantly evolving, and they continuously take on new tasks. The role of administrative professionals allows scientists to keep their main focus on the science and be less distracted by necessary administrative operations. Research administrators are a valuable asset and deserve professional recognition.

The mantra of many administrative professionals is "No job is too big or too small." They contribute greatly to the success of research and have a strong impact on science. As Kim C. Carter, SRAI president, says, "In the world of ordinary mortals, research administrators are superheroes" [16].

## Acknowledgments

Romina Garrido is assistant to the Bioinformatics and Genomics program of the Centre for Genomic Regulation (CRG), coordinated by Roderic Guigó, PhD. Nana Tamura is assistant for Piero Carninci, PhD, Center for Integrative Medical Sciences at RIKEN. Casandra A. Trowbridge is an administrative and operational director in the lab of Michael P. Snyder, PhD in the Department of Genetics at the Stanford School of Medicine at Stanford University. The authors want to thank Roderic, Piero, and Mike for their unconditional support and for creating the right environment for writing this article. The authors would also like to thank Kristin Ardlie, PhD at the Broad Institute of MIT and Harvard, who also played a key role with Roderic Guigó in setting collaborations between two of the authors.

Romina Garrido wants to thank the CRG for supporting her visit to Stanford in 2016 and to the RIKEN in 2018. Both visits have helped to lay the groundwork for writing these ten simple rules.

The authors want to thank Rut Carbonell and Bruna Vives for their help in reviewing the text.

## References

1. Nature (Ed.) (2018). Research institutions must put the health of labs first. *Nature*. 2018 May; 557 (7705):279–280. https://doi.org/10.1038/d41586-018-05159-0 Epub 2018 May 16. PMID: 29769683
2. Van Noorden R. (2018). Some hard numbers on science's leadership problems. *Nature*. 2018 May; 557 (7705):294–296. https://doi.org/10.1038/d41586-018-05143-8 Epub 2018 May 16. PMID: 29769686
3. Gillard S. (2009). Soft Skills and Technical Expertise of Effective Project Managers. *Issues in Informing Science and Information Technology*. 2009; 6:723–729. Available from: http://iisit.org/Vol6/IISITv6p723-729Gillard599.pdf.
4. SkillsYouNeed. (2016). Ways to Develop Soft Skills [web post]. [cited 2018 June 16]. https://www.skillsyouneed.com/rhubarb/develop-soft-skills.html.
5. Kerridge, Simon, Scott, Stephanie F. (2018) Research Administration around the World. Research Management Review, 23(1). pp. 1–34. ISSN 1068-4867. [2019 Mar 29]. Available from: https://kar.kent.ac.uk/68543/ Schnoebelen, N. (2018).
6. Critical Thinking in the Decision-Making Process. American Society of Administrative Professionals (ASAP). [cited 2019 June 16]. https://www.asaporg.com/critical-thinking-in-the-decision-making-process.

7. Coccia M. (2009). Bureaucratization in Public Research Institutions. *Minerva*. 2009 Mar; 47(1):31–50. https://doi.org/10.1007/s11024-008-9113-z Published online 2009 Feb 24. PMID: 19816535
8. Altmetric. Altmetric [website displaying tool] [cited 2019 Jun 17]. https://www.altmetric.com/.
9. Harris, N. (2009). University Administrator—Got What It Takes?. [cited 2009 Jan]. https://www.jobs.ac.uk/careers-advice/working-in-higher-education/1249/university-administrator-got-what-it-takes.
10. Kerridge S. (2019). Is there a need for formal Job qualifications in RMA in Europe? Professional recognition for research managers and administrators (RMAs). 2019 March; 2019 EARMA Conference. https://www.earma.org/wp-content/uploads/2019/05/ARMA-PDF-etc-for-EARMA-2019-final.pdf.
11. Kerridge S., & Scott S.F. (2018). Research Administration around the World. *Research Management Review*. 2018; 23(1). Available from: http://www.ncura.edu/Portals/0/Docs/RMR/2018/v23_n_1_Kerridge_Scott.pdf.
12. Enfield K.B., & Truwit J.D. (2008). The purpose, composition, and function of an institutional review board: balancing priorities. *Respir Care*. 2008 Oct; 53(10):1330–6. PMID: 18811996.
13. Kwok R. (2018). How to fit in when you join a lab abroad. *Nature*. 2018 May 23; 557, 599–601. https://doi.org/10.1038/d41586-018-05215-9 Epub 2018 May 23. PMID: 29789744
14. Trompenaars F., & Hampden-Turner C. (2012). Riding the Waves of Culture: Understanding Diversity in Global Business, 3rd Ed. MacGraw-Hill Education. 2012 Jan 10.
15. Halder, M., Binder, J., Stiller, J., Gregson, M. (2007). An Overview of the Challenges faced during Cross-Cultural Research. Nottingham Trent University. https://www.nottingham.ac.uk/sociology/documents/enquire/volume-8.pdf.
16. Carter, K. (2019). View from the Top. In the World of Ordinary Mortals, Research Administrators Are Superheroes [SRAI Blog Post] [cited 2019 Jun 17]. https://www.srainternational.org/blogs/srai-news/2019/05/02/view-from-the-top-in-the-world-of-ordinary-mortals?utm_source=SRA+International+Communications&utm_campaign=3673778b02-MAY_CATALYST_NONMEMBERS_2019_05_03&utm_medium=email&utm_term=0_dc20401f01-3673778b02-520720629.

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules to initiate and run a postdoctoral association

**Chiara Bruckmann[1,2]*, Endre Sebestyén[1,2]***

**1** IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy, **2** IPA (IFOM Postdoc Association), IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy

* chiara.bruckmann@ifom.eu (CB); endre.sebestyen@ifom.eu (ES)

## Introduction

Postdoctoral researchers (postdocs) are individuals who have obtained a PhD degree and are working in a temporary, mentored research position [1,2]. A proper postdoctoral experience should provide the necessary training for these individuals to achieve scientific and professional independence. During this phase of researchers' careers, they should develop new skill-sets by working in research groups. At the same time, they are usually involved in multiple tasks, such as mentoring trainees, applying for grants, writing publications, and keeping up with the latest advances in their field [3,4]. As young researchers enter the late-PhD and post-PhD period, they apply for positions with mentors whose work interests them. Mentors help postdocs prepare to lead research groups and improve existing skills or develop new ones [5].

In science, the number of researchers holding a doctoral degree and looking for postdoc positions has doubled in less than 2 decades [4]. However, the number of new academic jobs has not kept pace with this increase. Despite their restricted career prospects, stemming from the highly competitive environment they face [4,6–9], most postdocs wish to remain in research, and nonacademic career paths are often seen as failure [8,10–14]. At the same time, specific training that emphasizes lab management or transferable skills, opportunities to transition into alternative career paths, and coaching on complementary skills are scarce [15,16]. As early-stage researchers are often encouraged to broaden their experience through a stay abroad [2], a postdoctoral position is frequently accompanied by a geographical transition, along with all the challenges of living in a different country and leaving circles of friends and networks of colleagues [17,18].

Postdoctoral associations can improve the quality of the postdoctoral experience and provide support for professional development and everyday working life. Unlike graduate students, postdocs do not have classmates to whom they can turn for support and networking; therefore, a structure that facilitates their integration into their institution can instantly improve their quality of life. As postdocs pay neither tuition nor student fees, institutions dedicate substantially fewer resources to them, even though they constitute a large fraction of the labor force and, together with PhD students, are the major engine driving research. Therefore, a postdoc association can contribute significantly to the professional development of postdoctoral researchers by organizing networking and career events, thus widening their prospects [19]. An established postdoctoral association can facilitate open communication with the host institution and administrative bodies while also helping to develop a nurturing research environment. Additionally, active participation in an association will help postdocs prepare for leadership roles outside the lab and strengthen their curricula vitae (CVs) [20]. At present, only a small number of postdocs spend time developing additional skills beyond research skills during their postdoctoral period.

Check for updates

OPEN ACCESS

**Citation:** Bruckmann C, Sebestyén E (2017) Ten simple rules to initiate and run a postdoctoral association. PLoS Comput Biol 13(8): e1005664. https://doi.org/10.1371/journal.pcbi.1005664

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** August 17, 2017



**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

A postdoctoral association is beneficial not only for the postdocs but also for the hosting institution, as clear communication and representation of postdoc problems or needs leads to a better work environment. Top-level research is just a part of the puzzle that attracts talent to an institution. A positive work environment is self-reinforcing: open communication and internal networking support one another, and together, they create an environment that is creative, fun, hardworking, and highly productive [21].

In the United States, postdoctoral associations are present in many research institutes and universities, and the National Postdoctoral Association (NPA) organizes an annual conference, besides providing resources for postdocs and (where present) for postdoctoral offices [1]. However, there is little central coordination of different local associations. In the rest of the world, postdoc associations are sporadic and even more isolated than in the US. By sharing our experience in starting and running a postdoc association, and the challenges and joys we encountered, we wish to encourage postdocs to start other postdoc associations.

To start and run a postdoctoral association, we suggest the following 10 basic steps.

## Rule 1: Recruit a handful of motivated postdocs interested in starting a postdoc association

Call a meeting and advertise it by putting up flyers throughout research buildings, seminar rooms, libraries, and cafeterias and by talking directly to your own colleagues. Expect only a handful of attendees—but those who will show up will most likely express interest in setting up a postdoctoral association. Define the general aims of the association. The initial mission of a postdoc association can be very generic at first—for example, to enhance scientific interactions among postdocs and improve the quality of their personal life. Later, it can be tailored for the specific needs of your institution.

## Rule 2: Work together with other entities within the institution, such as human resources and senior administration

Links with administrators in graduate and postdoctoral offices, human resources, communications, grant management, and international offices are fundamental. It is important to facilitate discussions between postdocs and other entities within the institution in order to make the community more inclusive and integrated and to encourage collaboration and sharing of ideas. The postdoc association can work together with administrators, for example, to gather data on postdoc needs, keep an up-to-date email list of all postdocs, establish an alumni database of graduates who are working scientists, and set up a common calendar of events. This facilitates internal institutional networking and promotes personal and scientific relationships between present and past researchers of the institution [22]. Ideally, every institution should create a postdoctoral service office or similar unit (run by permanent employees) to keep postdoctoral researchers engaged in planning their careers and foster programs for their benefit [23]. The dearth of such postdoctoral offices and representatives is a significant reason why the challenges and problems of postdoc life remain practically invisible. If creating a postdoctoral office is not feasible (for example, if an institution does not have enough postdocs), the institution could consider giving the responsibility of supervising postdoctoral researchers to an existing administrative office—for example, to the graduate office. Participation in university committees and regular meetings with the relevant dean or the institution's chancellor or director of the institution are important opportunities for discussing postdoc issues and representing postdoc interests.

## Rule 3: Call for official elections of the association board and write statutes

Approximately 5–10 representatives, meeting monthly, should constitute the board of the association. The representatives should be elected by postdoctoral researchers themselves, and their mandate should be annual, as postdocs are often hired for 1 or 2 years. Nominees should include postdocs from multiple departments in order to guarantee diversity and the expression of various interests. After its election, the board should elect a president and a vice-president. Besides these 2 roles, in our experience, it might be difficult to define a specific structure for young associations. In the beginning, roles and responsibilities should be flexible, as postdocs come and go and also need to deal with their own deadlines and core responsibilities as researchers. When the association is more mature and well established within the institution, and when successful generational turnover of the postdoctoral representatives is therefore supported, you can define clearer structures, procedures, and roles for officers or board members [24]. As the workload for active postdoctoral representatives can be heavy, even when convening to address administrative matters, it is advisable to try avoiding meetings that last for more than 1 hour so that people are able to return to their other pressing duties. If the institution has a high percentage of international postdoctoral fellows, the board should be required to encompass representatives from different nationalities, thus gathering diverse views and opinions. Bear in mind that even an experienced postdoctoral association will face problems with issues such as the recruitment of new representatives and overcoming generational change, and this is particularly true for a newborn association. A successful strategy for volunteer recruitment should include extensive advertisement (both physical and electronic, for example, via a postdoc email list), active networking at social events and initiatives, and practical support on advertisement from administrative and human resource offices. The statutes of the postdoctoral association should be clear, concise, and distributed together with the aims of the association, by the human resource office to every new postdoc joining the institution, as part of a welcome package.

## Rule 4: Listen to the needs of your colleagues by running regular surveys and organize events to respond to different needs

As a first initiative of the newly founded postdoctoral association, organize a short meeting, inviting all postdocs, and distribute an anonymous questionnaire to help you understand their needs in your institution. The survey could include detailed questions on gender, dependents living with postdocs, satisfaction with salaries or working conditions, and long-term career goals (inter alia). All of this information will help in planning more useful events and understanding postdoc life generally. One's colleagues may not just want to get classic scientific training but may be eager to learn complementary skills, create connections to local companies, or just put together sport tournaments among colleagues. As a general rule, tailor the location, timing, and content of events to postdoc needs so that they are motivated to attend. For example, these events should be organized on or near campus, around lunchtime, at the end of working hours, or at another convenient time; this way, people can show up for an hour and then return to work or go home.

## Rule 5: Provide support for international postdocs

Many researchers move abroad for their postdoctoral research; therefore, they can feel out of place and isolated in an unfamiliar culture and can find it difficult to integrate socially as they spend most of their time working (a vicious cycle). If your institution lacks an international or diversity office, provide support for international postdocs yourselves and ensure that their particular needs are addressed. This means connecting with administrative and human resource personnel who can help in integrating newcomers. A postdoctoral association can support international postdocs by offering international lunches or coffee hours or by promoting the culture of the hosting country, for example, by organizing visits to exhibitions or restaurants. Together with the administration, the association should also provide language courses. Institutions that ensure social integration of postdoctoral fellows are more easily able to attract talent from abroad, which means this should also be a priority from an institutional perspective. A description of the association should be present on the institution's website, together with information on past events.

## Rule 6: Organize and promote career development events

Organize career events and professional development courses for postdocs and help to educate postdocs on the transferable skills they have that will be useful for promoting their careers both inside and outside academia. Various companies run tailored workshops for transferable skill development or project management. Unfortunately, not all institutions can afford to outsource this kind of training; therefore, while looking for an external sponsor (see Rule 10), the postdoc association, together with the postdoctoral office, can take advantage of internal know-how. For example, you could recruit principal investigators for mock interview workshops in which they explain the application path for academic positions and give tips and suggestions for successful proposals, CVs, and interviews [25]. In addition, you could invite successful alumni pursuing careers outside of classic academic research to share their personal career paths and lessons. Many postdocs do not know where to find nonacademic career options or how to prepare for them. People who have experienced such situations can advise on choosing next career steps, considering not only the responsibilities of positions and postdocs' desired salary ranges but also their desired lifestyles. These meetings enable postdocs to talk about their career plans, do professional development activities, and ideally, network with scientists who have completed a PhD but moved outside of academia into the domains of industry, scientific publishing, technical sales, start-ups, teaching and academic support, and so on.

## Rule 7: Get visibility in your institution, using social networks, newsletters, and flyers

Every postdoctoral researcher joining the institution should automatically become part of the association and the postdoc mailing list and thus should be regularly informed about ongoing events and programs. In addition to the mailing list, the association should promote events using a periodic newsletter and flyers. If possible, there should be dedicated spaces in high-traffic areas for posters, flyers, and information on the association. A dedicated website, linked to social media such as LinkedIn, Facebook, and Twitter, should include a short description of the aims of the association, the profiles of the elected representatives, and past and upcoming events. This can be used by the association to disseminate information to current and past postdocs and additionally gives geographically unrestricted visibility, facilitating networking with other associations. Furthermore, postdocs considering moving to the host institution can get in touch through social media or the website and get a sense of the local community.

## Rule 8: Build a community by organizing social events

Balance scientific events with social events. Encourage networking and exchange of ideas by organizing monthly meet-up events for postdocs in an informal setting, such as dinners out, visits to exhibitions/museums, seasonal barbecues, or sport groups (see also Rule 5). These activities can create a supportive peer group and integrate postdocs, helping members socialize, make friends, and discuss work or career steps.

## Rule 9: Network with other postdoctoral associations at the national and international levels

In the US, postdoctoral associations are well established in many institutions and give a voice to the postdoctoral community. These associations are supported by the NPA if they apply for membership. In the rest of the world, postdoctoral associations are not coordinated at any level and usually work in isolation; however, even regional coordination between associations can increase the visibility of the postdoctoral community and facilitate access to training, networks, and events. Some examples include the Postdoctoral Association of the European Molecular Biology Laboratory (EMBL), which is active in 5 cities and 4 countries [26], or the Marie Curie Fellowship Association [27]. However, the former is still limited to 1 institute, while the latter is restricted to Marie Curie Fellowship winners. Connections with other associations may serve as the basis for shared projects such as joint career events, international symposia, and joint funding applications.

## Rule 10: Get sponsored by companies

In order to support large events, you could try to attract a broad range of sponsors and donors such as vendors or industry professionals in different fields of science. For example, you can offer different sponsorship packages, agree to distribute products or company-specific information to event participants, place company logos and links on event websites, or set up promotional booths in event exhibition areas (although you should first check institutional policies, which might limit promotional opportunities or the types of events that can be funded). Vendor fairs and scientific talks given by company representatives are some examples of sponsored events that postdoc associations can use for fundraising. By reaching out directly to participants, sponsoring companies will benefit through advertisement and increased visibility, while postdocs will benefit from the opportunity to hear highly qualified speakers. Group leaders or alumni members who are moving or have moved to industry can provide an extended network of contacts that the association can approach. Companies might also be willing to provide courses by involving their internal human resources and training teams in exchange for the visibility and networking they will garner.

## Conclusions

Postdoctoral researchers are often isolated, their careers at the mercy of their principal investigators or advisors. Unlike students, their affiliation is often not to a department but rather to a specific lab. Establishing a postdoctoral association can be a way to foster networks needed for effective support; however, setting up and running a postdoctoral association requires time commitment from active members and so should ideally be promoted by the institution as part of the postdoc experience.

The need for networking and support among postdocs is strong, but getting people involved in initiating a postdoctoral association might nevertheless be seen as a major obstacle. However, based on our experience, we can say that it is feasible and rewarding. Our hope is that after reading this article, many postdoctoral researchers will feel empowered to start a postdoctoral association at their home institutions. Given the temporary nature of the postdoc, it might take multiple generations to get an association running at full speed, but this should not discourage its founders from taking the initiative and making this important contribution.

A positive postdoctoral experience should include establishment of a solid network of colleagues. Accordingly, institutions should motivate postdoctoral fellows to stay in touch with one another and to continuously expand their networks in the present and future.

Finally, we encourage existing postdoctoral associations to network at the national and international levels, taking as an example the NPA in the US. There are a few independent postdoc associations in the European Union that would greatly benefit from coordination with the aim of setting up scientific meetings, networking, applying for common funding for advanced training, and gathering information on postdoctoral lives, career prospects, and needs. As there is no census of existing postdoctoral associations in Europe, and only a small number of reports discuss working conditions, salaries, gender inequality, training, or institutional support, networking would yield excellent opportunities to start collecting this information.

The IFOM (Fondazione Italiana per la Ricerca sul Cancro [FIRC] Institute of Molecular Oncology) Postdoc Association is a newborn association based at IFOM, Milan (Italy), and it would be exciting to see it grow through networking with other Italian and European postdoctoral associations.

## Acknowledgments

The authors of this article met through the IFOM Postdoctoral Association (IPA, pdassociation-team@ifom.eu). We would like to acknowledge our colleagues who served as board members in 2016: Sina Atashpaz, Christopher Bruhn, Carmen Livi, Chiara Malinverno, Umberto Restuccia, and Maria Vinciguerra. We are in debt to Elena Bauer, Mio Sumie, and Marina Properzi for administrative support. The establishment of IPA was encouraged and supported by Marco Foiani, the IFOM scientific director, and the founding postdoctoral fellows in 2013 were Yathish Achar, Salvatore Cortellino, Audrey Laurent, Hiroaki Kajiho, Gema Malet, Muege Ogrunc, Kerstin Schmitz, Barnabas Szakal, Gabriele Varano, and Hans-Peter Wollscheid. We are grateful to Francesco Blasi for his critical reading of the manuscript and to Kristina Havas-Cavalletti for stimulating discussion.

## References

1. National Postdoctoral Association. In: nationalpostdoc.org [Internet]. [cited 26 Jun 2017] http://www.nationalpostdoc.org/
2. Dillon N. The postdoctoral system under the spotlight. EMBO Rep. 2003; 4: 2–4. https://doi.org/10.1038/sj.embor.embor721 PMID: 12524507
3. Cain B, Budke JM, Wood KJ, Sweeney NT, Schwessinger B. How postdocs benefit from building a union. eLife. 2014; 3: e05614. https://doi.org/10.7554/eLife.05614 PMID: 25415240
4. Maher B, Sureda Anfres M. Young scientists under pressure: what the data show. Nature. 2016; 538: 444. https://doi.org/10.1038/538444a PMID: 27786225
5. de Ridder J, Meysman P, Oluwagbemi O, Abeel T. Soft Skills: An Important Asset Acquired from Organizing Regional Student Group Activities. PLoS Comput Biol. 2014; 10: e1003708. https://doi.org/10.1371/journal.pcbi.1003708 PMID: 24992198
6. National Research Council, Committee on Science, Engineering, and Public Policy, Committee to Review the State of Postdoctoral Experience in Scientists and Engineers (2014) The Postdoctoral Experience Revisited. The National Academies Press, Washington DC. https://www.nap.edu/catalog/18982/the-postdoctoral-experience-revisited
7. van der Weijden I, Teelken C, de Boer M, Drost M. Career satisfaction of postdoctoral researchers in relation to their expectations for the future. High Educ. 2016; 72: 25–40. https://doi.org/10.1007/s10734-015-9936-0
8. Powell K. The future of the postdoc. Nature. 2015; 520: 144–147. https://doi.org/10.1038/520144a PMID: 25855437

9. Should I become a professor? Success rate 3%! In: smartsciencecareer.com [Internet]. [cited 26 Jun 2017] http://www.smartsciencecareer.com/become-a-professor/
10. Harsh reality. Nature. 2014; 516: 7–8. https://doi.org/10.1038/516007b PMID: 25471843
11. Zambetti LP. Away from the bench: non-academic careers for a postdoc. PostDoc J. 2013; 3: 15–16.
12. The postdoc experience: hopes and fears. In: timeshighereducation.com [Internet]. [cited 26 Jun 2017] https://www.timeshighereducation.com/the-postdoc-experience-hopes-and-fears
13. Academia doesn't have a PhD problem, it has an attitude problem. In: contemplativemammoth.com [Internet]. [cited 26 Jun 2017] https://contemplativemammoth.com/2013/06/19/academia-doesnt-have-a-phd-problem-it-has-an-attitude-problem/
14. Dickey Eleanor. The impact of the poor academic job market on PhD graduates. CUCD Bulletin 2014; 43.
15. Rohn J. Give postdocs a career, not empty promises. Nature. 2011; 471: 7. https://doi.org/10.1038/471007a PMID: 21368781
16. Rybarczyk BJ, Lerea L, Whittington D, Dykstra L. Analysis of Postdoctoral Training Outcomes That Broaden Participation in Science Careers. CBE Life Sci Educ. 2016; 15. https://doi.org/10.1187/cbe.16-01-0032 PMID: 27543634
17. Kumar A. Adaptability in life and work. Science. 2016; 353: 954. https://doi.org/10.1126/science.353.6302.954 PMID: 27563097
18. Uitto J. The benefits of international postdoctoral research fellowships: a personal perspective. J Invest Dermatol. 2013; 133: 2301–2302. https://doi.org/10.1038/jid.2013.306 PMID: 24030641
19. Michaut M. Ten Simple Rules for Getting Involved in Your Scientific Community. PLoS Comput Biol. 2011; 7: e1002232. https://doi.org/10.1371/journal.pcbi.1002232 PMID: 22046114
20. Sweedler JV. Career Advice for Graduate Students and Postdoctoral Fellows. Anal Chem. 2016; 88: 2513–2514. https://doi.org/10.1021/acs.analchem.6b00532 PMID: 26863097
21. The 50 best companies to work for in America. In: businessinsider.com [Internet]. [cited 26 Jun 2017] http://www.businessinsider.com/payscale-best-companies-to-work-for-in-america-2016-4
22. Silva EA, Des Jarlais C, Lindstaedt B, Rotman E, Watkins ES. Tracking Career Outcomes for Postdoctoral Scholars: A Call to Action. PLoS Biol. 2016; 14: e1002458. https://doi.org/10.1371/journal.pbio.1002458 PMID: 27152650
23. Ross RG, Greco-Sanders L, Laudenslager M. An Institutional Postdoctoral Research Training Program: Increasing Productivity of Postdoctoral Trainees. Acad Psychiatry. 2016; 40: 207–212. https://doi.org/10.1007/s40596-015-0281-5 PMID: 25876090
24. CONSTITUTION & BYLAWS—MIT Postdoctoral Association. In: mit.edu [Internet]. [cited 26 Jun 2017] https://pda.mit.edu/about/bylaws/
25. Henderson RI, Syed N. The Mock Academic Faculty Position Competition: A Pilot Professional and Career Development Opportunity for Postdoctoral Fellows. Acad Med. 2016; https://doi.org/10.1097/ACM.0000000000001111 PMID: 26862841
26. Postdoctoral Programme—Postdoctoral Association—EMBL. In: embl.de [Internet]. [cited 26 Jun 2017] https://www.embl.de/training/postdocs/12-postdoc_assoc/index.html
27. Marie Curie Alumni Association. In: mariecuriealumni.eu [Internet]. [cited 26 Jun 2017] https://www.mariecuriealumni.eu/

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten simple rules for forming a scientific professional society

**Bruno A. Gaëta[1]*, Javier De Las Rivas[2], Paul Horton[3], Pieter Meysman[4,5], Nicola Mulder[6], Paolo Romano[7], Lonnie Welch[8]**

**1** School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia, **2** Bioinformatics and Functional Genomics Group, Cancer Research Center (CiC-IBMCC, CSIC/USAL/IBSAL), Salamanca, Spain, **3** Artificial Intelligence Research Center, AIST, Tokyo, Japan, **4** Biomedical Informatics Research Center Antwerp (biomina), University of Antwerp/Antwerp University Hospital, Antwerp, Belgium, **5** Department of Mathematics and Computer Sciences, University of Antwerp, Antwerp, Belgium, **6** Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, **7** Bioinformatics group, Proteomics laboratory, IRCCS AOU San Martino IST, Genoa, Italy, **8** School of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio, United States of America

* bgaeta@unsw.edu.au



OPEN ACCESS

**Citation:** Gaëta BA, De Las Rivas J, Horton P, Meysman P, Mulder N, Romano P, et al. (2017) Ten simple rules for forming a scientific professional society. PLoS Comput Biol 13(3): e1005226. https://doi.org/10.1371/journal.pcbi.1005226

**Published:** March 23, 2017



**Funding:** The authors received no specific funding for this work.

**Competing interests:** I have read the journal's policy and have the following conflicts: all the authors are members and/or officers of the International Society for Computational Biology (ISCB), which is mentioned in the manuscript as an example of a professional society with which to partner when starting a computational biology professional society.

In this article, we present ten simple rules to assist in the formation and management of a scientific professional society. We define a "scientific professional society" (also often referred to as a "scholarly society") as an association of people who come together to promote progress in a specific technological or scientific area and to facilitate the interaction of interested people on a regional, national, or international level. For the rest of this article, we will use the term "society" to refer to a scientific professional society.

Some emerging disciplines in biology where scientific professional societies have appeared relatively recently include computational biology and bioinformatics, systems biology, and functional genomics. The main professional body representing computational biology is the International Society for Computational Biology (ISCB) [1], founded in 1997. For systems biology, the International Society for Systems Biology (ISSB) [2] started as an International Conference created by Hiroaki Kitano in 2000, which led to the formation of the society. In the field of functional genomics, the Functional Genomics Data Society (FGED) [3] was formed in 1999 as the Microarray Gene Expression Data Society (MGED) in recognition of the need to establish standards for sharing and storing genomic data. MGED changed its name in 2010 to FGED. These examples show that the formation of a new scientific professional society is often triggered by the emergence of a new scientific discipline or by technological developments in science. This, of course, is not the only reason for developing a new society. Some computational biology-related societies with a very targeted focus include the International Society for Biocuration (ISB) [4,5], which is aimed at ensuring the best possible annotation methods and tools for curation of biological databases, and the Global Organization for Bioinformatics Learning, Education and Training (GOBLET) [6,7], whose aim is the development of the best possible educational strategies and materials in bioinformatics.

However, the majority of societies aim to foster networking in a narrower geographical area, such as a city or a country, and these are the focus of this article. The rules presented draw on the collective experience of the authors, who all have been active in the formation and/or development of computational biology societies in their respective regions and countries in Asia [8,9], Africa [10], North America [11], Europe [12], and Ibero-America [13].

These “rules” are by no means prescriptive but are meant to act as guidelines and suggestions to help avoid some common pitfalls.

## Rule 1: Have a clear scope and vision

Setting up a new society will require a lot of work—it is therefore a good idea, before getting started, to identify the goals and the target audience of your society, as these will have a strong impact on the process. A society’s goals can include improving the profile of a scientific discipline through public advocacy and government representation, and fostering networking, information sharing, mentoring, career opportunities, leadership training, and professional development. To couch this process in strategic management terms [14], identifying the "vision" of the society (its purpose and aspirations) is important to guide your "strategy" (the practical steps to achieve the vision), which in turn determines the processes and resources required. For example, a society aiming to run a large annual conference will require setting up significant infrastructure to manage budget and liability, in contrast to a society whose goal is to run a monthly seminar series in a local university and can get by with a mailing list and a small cash pool for drinks and snacks.

The questions you need to ask yourself at this stage are similar to those a new business needs to address: what "products" or "services" are you going to offer? These could include conferences, seminars, publications, networking events, student scholarships, and public advocacy. Who are your intended customers in terms of scientific discipline, geographic region, and career stage? What “benefits” is your society going to provide to these members? Who is your “competition”: other local societies or larger regional/national/international societies covering similar or related scientific domains? And how will your proposed society differentiate itself from these so that it will attract and retain members? If other societies exist in related scientific areas, you need to make sure you offer something unique, but at the same time, look for opportunities to dynamically partner with other societies, for example for joint meetings (see Rules 2 and 9). While you do not want to be delayed by over-planning, it is important to think about these big questions in order to start on the right foot and ensure your resources are spent productively and, more importantly, that you really want to commit your time and energy to a new society instead of an already existing group [15].

The answers to these questions will form the business plan of your society, and like any business plan, they will need to be revisited at regular intervals (for example, every two years) as the society and its landscape evolve.

## Rule 2: Start an annual meeting

An annual gathering, such as a conference [16] or a workshop [17,18], will propel you forward as you build a community of individuals with a common interest. The meeting will provide a focus for your interactions and will provide the opportunity to interact regularly throughout the year as you plan the event. An annual in-person meeting provides the opportunity to build relationships and to have spontaneous ideas for enhancing the offerings of your community.

When planning the meeting, form an organizing committee that spans the geographic region and that includes diverse perspectives (e.g., public and private universities, government research facilities, medical research labs, and industry). Encourage members to seek buy-in from their home institutions in the form of co-sponsorship for the event. Among the invited keynote speakers, include some people from the local region. If possible, try to partner with an international society in the same area (for example, ISCB for computational biology) to tap into their meeting organization expertise and to help with event promotion. Finally, start small, and make incremental improvements to the meeting each year.

## Rule 3: Make history: Document everything

You may have just started, but the years will go by quickly, and (documented) tradition lends credibility. So spend a moment early on to consider how you will document the activity of your society and the contribution of your members to your society and the research community at large. Plan to archive the webpages of all of your society's major events, especially your annual meetings, starting from the first one! A few years down the road, consider starting some awards, which may grow to become prestigious in the future.

Internal documentation is also essential, both for transparency of decision-making and succession planning as the society changes over time and new officers come on board. Archive all meeting minutes and, more importantly, all contact information, including that of officers and sponsors, using a collaborative online platform that can be accessed readily by current and new society officers.

## Rule 4: Start small or build on existing scientific and professional networks

A professional society cannot be fully designed by a single scientist alone: it requires consensus, as well as exchange of ideas, expertise, and skills. Required skills are very different from research expertise and, without suitable mentoring, can only be acquired by (many) trials and (some) errors. If at all possible, try to take advantage of the experience available in existing societies and networks in your area. The benefits are both practical (for example, lists of interested parties and service providers) and conceptual, including ways of sharing ideas and finding consensus.

In most cases, starting with a small number of resolute members is more efficient than trying to be all-encompassing from the start. It makes it easier to collaborate and share a common vision, assign duties, and make consensus decisions. Starting small, however, does not mean that you should remain small. A strategy for growing regularly by adding new members or involving more institutions must also be defined (for example, through scientific events that grow larger from year to year, involve a range of participants, and are located where potential members are).

## Rule 5: Set up your governance structure and constitution

Formalising the structure of the society is an essential step towards a working association. The task may seem daunting at first, so it is best to start with a simple set of terms of reference and by-laws and grow from there as the society expands. You may want to seek inspiration in the constitution and articles of governance of existing societies, many of which are directly available from their website: see, for example, the by-laws of ISCB [19] or of the American Society for Human Genetics [20]. One generally successful approach is to choose an initial executive committee made up of committed volunteers, with, if at all possible, an emphasis on inclusivity and diversity, including geographic and gender diversity. A typical executive committee structure includes a president, a secretary whose role is to keep records and minutes, a treasurer who focuses on the all-important financial aspects, and a few vice presidents with a focus on specific missions of the society (education, conferences, communications, industry outreach). A smaller group is more efficient; however, it is essential that the duties of the society are not left on the shoulders of one or two people.

The executive committee can then formulate an initial set of by-laws covering executive and committee terms and election procedures, as well as create relevant subcommittees and define their terms of reference. An emphasis on diversity, including gender and career stage diversity, is important at this stage to ensure the leadership is representative of its members. Organisation management is not a standard scientific skill, so don't plan to do everything yourself, and be ready to seek advice. As the society (and its finances) grow, hiring a dedicated executive professional may become necessary.

## Rule 6: Be inclusive

Starting a society with some of your closest contacts and frequent collaborators is easy but can limit the growth potential of your new group, as you will tend to have similar contacts and travel in the same circles. It is therefore critical to think beyond your partners in research and reach out to those whom you might not have had reason to contact before. Try not to exclude anyone who wants to get involved, especially early on. Furthermore, especially for organisations with a regional focus, it is important to be inclusive and span the breadth of your domain at the start. Your outreach will be greatly aided by including individuals from different universities and institutions. For example, each institution has their own procedures for mass mailing, and some might have valuable funding or sponsorship programs. Having someone on the inside can help to navigate these procedures. Reaching out to these people can be as simple as talking to them during conferences and discussing your plans during one of the breaks. Sending a personalised email or making a call to potential partners about your intentions to form a new society can sometimes yield surprising results. If they themselves might not be interested, they often know someone who might be. Inclusiveness is also essential for accountability. The society should not be perceived by outsiders as a closed shop or an exclusive club but should strive to be open and act for the benefit of the community, rather than a select few.

Once the society is up and running, it is critical for its longevity that new people can still become involved easily. A web form or contact email on the website may seem impersonal, but it can be very useful to direct interested parties. Speak to people at different events and get them excited about the planned activities and the society as a whole. In addition, always be on the lookout for individuals who might be affiliated with institutions that are currently not represented.

Many successful societies have a student-focused component aimed at PhD students and other junior scientists, often run by the students themselves [21]. This can take the form of local student-run chapters, dedicated student seminars/workshops, and even local journal clubs [22]. Creation of a solid support structure, along with finding a motivated group of students to get things started, is critical for the development of such student-run activities. Simply providing access to the infrastructure that has been set up for the society, such as mailing lists, web space, or even bank accounts, will facilitate their start-up. It is important to not just assign tasks to the students, however tempting it may be to use them as cheap labor, but to let them develop their own plans and activities within the society: they will be more motivated to help out, and this will increase their sense of belonging to the group. In addition, for many, this is a very useful learning experience with regards to many different kinds of "soft skills" (such as organisation, communication, and leadership) that are often missing from a typical PhD program [23]. The students of the society may one day be its leaders, so it is important to show what such a society can do and mean for them from the start and to include student representatives in the leadership of the main body of the society.

## Rule 7: Plan the financial and legal aspects

If you want your society to be more than an informal, volunteer-run effort, it is going to require funding. The most common funding sources for societies are membership fees and conference proceeds, but you should also investigate what grants and funding schemes are available in your region that your society could leverage. In many countries, state, institutional, and philanthropic grants are available for purposes such as improving communication and networking between scientists, strengthening a particular scientific discipline, or meeting professional education needs. Such grants can provide a great source of funding for one-off projects or for seeding a new initiative and are often easier to obtain than major research grants. They are, however, not always well-publicised, and it's a good idea to set up a fundraising taskforce with members from a wide range of institutions that keeps an eye out for funding opportunities. Diversity is key, as taskforce members from different circles will be more likely to hear about grants and sponsorship opportunities from different sources. The key to many succesful societies is their ability to leverage grants supporting conferences and workshops [24,25] and networking/connection [26,27] for such purposes.

However, these grants are typically one-off, and more sustainable sources of funding are also necessary to provide for longer term planning and continuing growth. In this context, membership fees can be a double-edged sword, as they require a whole infrastructure to collect and administer, and more importantly, they will act as a deterrent for potential members. You should consider carefully whether the income from membership fees is absolutely necessary and whether your society offers sufficient value for its fees to attract and retain members.

Conferences, on the other hand, can be both a great way to meet some of your society's goals and a way to raise some income, but they do require a lot of commitment and work (see Rule 2). They also require a substantial starting investment to get underway.

In any case, if your society is going to collect and manage money, it will need an appropriate legal structure that will provide confidence to its members, as well as protect its officers from financial and legal liability. The exact requirements for this vary from country to country, and you need to familiarise yourself with your local laws and structures. Some countries will require the society to become incorporated as a nonprofit organisation, while others may allow a larger "parent" institution to manage the accounts to keep administrative overheads to a minimum and provide some legal "umbrella." For example, GOBLET and the European Molecular Biology network (EMBnet) [28] make use of the Dutch "Stichting" (private foundation) legal structure, while ISCB is incorporated as a nonprofit, tax-exempt corporation in the state of Delaware. These legal details can be difficult to navigate for scientists; however, many institutions have an in-house legal department that can provide guidance and some valuable advice. If your society is going to span multiple countries/states, it is also worthwhile to investigate which of these countries/states offers the most convenient laws and taxation regimes for your purposes.

## Rule 8: Design a clear communication strategy

Communication is key to a successful society. Members need to feel like they are part of an active community and kept informed of relevant activities. A professional society website is important as the main source of information on the society's aims, governance, events, and membership information. A yearly event calendar is a relatively low-maintenance resource that can draw existing and potential members to the website and advertise the society's events. The society should decide whether these sites will be open to the public or restricted to its members. Technically, there are a number of frameworks on which to build a professional-looking website that also allow you to keep track of your membership. These include the open source Drupal [29] and Wordpress [30] content management systems, which are nowadays preinstalled on many web hosts. A mailing list is essential for reaching all members, and one should have a mechanism in place to keep the list up to date as memberships lapse or are renewed. This list should be active with regular summaries of activities to show that the society is alive and keen to involve its members. Joining a society and then not hearing from them until renewal time does not provide an incentive to renew.

These formal mechanisms need to be complemented by an appropriate social media strategy that allows not only the society to communicate with members but also members to communicate with each other. Excellent strategies for the development of your online outreach have been proposed by Bik et al. [31] The goal should be to have a presence on the sites where the members are, which can vary from country to country. Facebook [32] has the highest user base; however, for many users, it is restricted to personal rather than professional contacts. LinkedIn [33] and Yammer [34] are focused on professional networking but have smaller user bases. Twitter [35] is a convenient avenue for short communications and fostering a sense of community. Sites such as Meetup [36] and EventBrite [37] are also useful for building social networks based around events. A Meetup group can even be used as a simple membership database and mailing list manager if you are not collecting membership fees. However, as with the mailing list, new content must be added regularly in order for the site (and, by extension, the society) to be perceived as active. Remember that community participation on social media can sometimes backfire, and suitable posting policies and active moderation should be put in place to avoid inappropriate postings that could tarnish the society's image.

It is also important to develop your branding and some marketing materials, such as brochures and posters that can be displayed at conferences to inform members about the society and to attract new members. Hosting an exhibition stand at conferences attracts interest but may be expensive. An alternative is a poster within the conference poster session. You could involve younger members of the community in the development of the branding and of marketing materials (for example, through competitions for logo designs or free membership for the brochure designer). The tone of the brochure should reflect the audience for which it is intended (slightly less formal and more fun for students and academics versus formal and professional for attracting industry partners). Marketing can also be achieved through publications in other organisations' newsletters (for example, ISCB and EMBnet for computational biology) or through regular society newsletters. Whichever mechanism is used for communication, make sure it is active to keep the members engaged and interested.

Internal communication between society officers also needs to be considered. Regular face-to-face meetings tend to be most efficient, but if that is not possible, a number of free teleconference options are available [38,39], provided the teleconference involves only a few participants. Some officers may also have access to commercial solutions through their institutions. Websites such as Doodle [40] and timeanddate [41] are useful for setting up meeting times, especially when the participants are in different time zones. Remember to keep good records of all the meetings (Rule 3).

## Rule 9: Seek partnerships and affiliations with other societies

All scientific societies have a goal of improving networking and communication in their areas, and in that context, significant synergies can be generated by partnering with more established societies. Benefits will include access to expertise and, often, to services including conference organisation (see Rule 2) and legal (see Rule 7) and administrative support. Any society whose goals include advocacy for its scientific discipline will gain a lot of traction by being able to call on the strength of a larger group. Your society may also be able to access some specific funding schemes and grants through partnership with a larger organisation.

For example, within the discipline of computational biology, it is valuable to partner with ISCB, which has in place a comprehensive affiliation scheme to support and collaborate with other bioinformatics societies. Groups that serve a particular geographic region can become regional affiliates of ISCB, while topic-based groups can align with ISCB as communities of special interests (COSIs). ISCB offers assistance to groups that are in the formative stages, provides opportunities for associations to have a global impact, and rewards groups that make significant contributions to the broader community. Connection with an international society leads to enhanced opportunities for collaboration, provides access to additional sources of expertise, makes available additional resources, and infuses additional energy.

## Rule 10: Have fun: Enjoy science and make friends

Academic societies are a serious business that promote crucial scientific collaboration. But at the same time, they are social clubs supported by volunteer time. Your society will be more sustainable if the members not only find the society's activities meaningful but also find them to be enjoyable and special experiences sharing time with other colleagues. One easy but effective way to facilitate this is to make sure you leave some free time (for example, lunchtimes and evenings) when you schedule events to allow the participants to socialize. One of the greatest side benefits of a society is the opportunity to meet and exchange experiences with many people of different backgrounds, with returns that can go beyond the scientific. By holding a meeting (especially if it is in an interesting venue), your society will provide a catalyst for fun and productive networking and, occasionally, even the start of lifelong friendships.

## Summary

Starting a professional society is not something that should be entered into lightly: it requires work and dedication that can detract from your research projects and other career objectives [15]. It certainly should not be attempted on your own. But there are many potential benefits and rewards in terms of promoting the profile of your discipline (which, in turn, can affect your grant success), boosting your own profile, developing useful management and leadership skills, finding mentors, and forming essential contacts and partnerships, as science is becoming increasingly collaborative. A successful society will be a source of lifelong learning and new ideas, will open up career opportunities for students and investigators, and will provide a much stronger voice for your discipline than an isolated scientist.

## References

1. ISCB—International Society for Computational Biology. In: iscb.org [Internet]. [cited 7 Jun 2016]. https://www.iscb.org/
2. The International Society for Systems Biology. In: issb.org [Internet]. [cited 7 Jun 2016]. http://issb.org/
3. fged.org. http://fged.org
4. International Society for Biocuration. In: biocuration.org [Internet]. [cited 7 Jun 2016]. http://biocuration.org/
5. Bateman A. Curators of the world unite: the International Society of Biocuration. Bioinformatics. 2010; 26: 991. https://doi.org/10.1093/bioinformatics/btq101 PMID: 20305270
6. GOBLET | Global Organisation for Bioinformatics Learning, Education & Training. In: mygoblet.org [Internet]. [cited 7 Jun 2016]. http://www.mygoblet.org/
7. Attwood TK, Atwood TK, Bongcam-Rudloff E, Brazas ME, Corpas M, Gaudet P, et al. GOBLET: the Global Organisation for Bioinformatics Learning, Education and Training. PLoS Comput Biol. 2015; 11: e1004143. https://doi.org/10.1371/journal.pcbi.1004143 PMID: 25856076
8. >> Asia Pacific Bioinformatics Network <<. In: apbionet.org [Internet]. [cited 7 Jun 2016]. http://www.apbionet.org/
9. Japanese Society for Bioinformatics—JSBi ホーム. In: jsbi.org [Internet]. [cited 7 Jun 2016]. http://www.jsbi.org/

10. ASBCB—African Society for Bioinformatics and Computational Biology. In: asbcb.org [Internet]. [cited 7 Jun 2016]. http://www.asbcb.org/
11. GLBIO/CCBC 2016. In: iscb.org [Internet]. [cited 7 Jun 2016]. http://www.iscb.org/glbioccbc2016
12. :. BITS Bioinformatics Italian Society. In: bioinformatics.it [Internet]. [cited 7 Jun 2016]. http://bioinformatics.it/
13. Sociedad Iberoamericana de Bioinformática / Iberoamerican Society for Bioinformatics (SoIBio) |. In: soibio.org [Internet]. [cited 7 Jun 2016]. http://www.soibio.org/
14. Carpenter M, Erdogan B, Bauer T. Principles of Management. Flat World Knowledge. Inc. USA; 2009.
15. Michaut M. Ten simple rules for getting involved in your scientific community. PLoS Comput Biol. 2011; 7: e1002232. https://doi.org/10.1371/journal.pcbi.1002232 PMID: 22046114
16. Corpas M, Gehlenborg N, Janga SC, Bourne PE. Ten Simple Rules for Organizing a Scientific Meeting. PLoS Comput Biol. 2008; 4: e1000080. https://doi.org/10.1371/journal.pcbi.1000080 PMID: 18584020
17. McInerny GJ. Ten Simple Rules for Curating and Facilitating Small Workshops. PLoS Comput Biol. 2016; 12: e1004745. https://doi.org/10.1371/journal.pcbi.1004745 PMID: 27441642
18. Pavelin K, Pundir S, Cham JA. Ten simple rules for running interactive workshops. Bourne PE, editor. PLoS Comput Biol. 2014; 10: e1003485. https://doi.org/10.1371/journal.pcbi.1003485 PMID: 24586135
19. ISCB bylaws. In: iscb.org [Internet]. [cited 30 Sep 2016]. https://www.iscb.org/bylaws
20. ASHG bylaws. In: ashg.org [Internet]. [cited 30 Sep 2016]. http://www.ashg.org/pages/about_bylaws.shtml
21. Shanmugam AK, Macintyre G, Michaut M, Abeel T. Ten Simple Rules for Starting a Regional Student Group. PLoS Comput Biol. 2013; 9.
22. Lonsdale A, Penington JS, Rice T, Walker M, Dashnow H. Ten Simple Rules for a Bioinformatics Journal Club. PLoS Comput Biol. 2016; 12.
23. de Ridder J, Meysman P, Oluwagbemi O, Abeel T. Soft Skills: An Important Asset Acquired from Organizing Regional Student Group Activities. PLoS Comput Biol. 2014; 10: e1003708. https://doi.org/10.1371/journal.pcbi.1003708 PMID: 24992198
24. NIH Support for Scientific Conferences (R13 and U13). In: grants.nih.gov [Internet]. [cited 30 Sep 2016]. http://grants.nih.gov/grants/funding/r13/index.htm
25. Conference Sponsorship Program 2016 [Internet]. [cited 30 Sep 2016]. http://www.chiefscientist.nsw.gov.au/investing-in-science/conference-sponsorship-program
26. www.kgf.ch. In: kgf.ch [Internet]. [cited 11 Oct 2016]. http://www.kgf.ch
27. www.cyted.org. In: cyted.org [Internet]. [cited 12 Oct 2016]. http://www.cyted.org
28. EMBnet | Bioinformatics without borders. In: embnet.org [Internet]. [cited 7 Jun 2016]. http://www.embnet.org/
29. Drupal. In: drupal.org [Internet]. [cited 30 Sep 2016]. http://www.drupal.org
30. Wordpress. In: Wordpress [Internet]. [cited 11 Oct 2016]. https://wordpress.com/
31. Bik HM, Dove ADM, Goldstein MC, Helm RR, MacPherson R, Martini K, et al. Ten simple rules for effective online outreach. PLoS Comput Biol. 2015; 11: e1003906. https://doi.org/10.1371/journal.pcbi.1003906 PMID: 25879439
32. Facebook. In: facebook.com [Internet]. [cited 30 Sep 2016]. https://www.facebook.com/
33. LinkedIn. In: linkedin.com [Internet]. [cited 30 Sep 2016]. http://www.linkedin.com
34. Yammer. In: yammer.com [Internet]. [cited 30 Sep 2016]. http://www.yammer.com
35. Twitter. In: twitter.com [Internet]. [cited 30 Sep 2016]. http://twitter.com
36. We are what we do | Meetup [Internet]. [cited 30 Sep 2016]. https://www.meetup.com/
37. Eventbrite [Internet]. [cited 30 Sep 2016]. https://www.eventbrite.com
38. www.skype.com. In: skype.com [Internet]. [cited 30 Sep 2016]. http://www.skype.com
39. Google Hangouts [Internet]. [cited 30 Sep 2016]. https://hangouts.google.com
40. Doodle: easy scheduling [Internet]. [cited 30 Sep 2016]. http://doodle.com/
41. www.timeanddate.com. In: timeanddate.com [Internet]. [cited 30 Sep 2016].

PLOS | COMPUTATIONAL BIOLOGY

EDITORIAL

# Ten Simple Rules for a Bioinformatics Journal Club

**Andrew Lonsdale[1]*, Jocelyn Sietsma Penington[2], Timothy Rice[3], Michael Walker[4], Harriet Dashnow[4,5,6]**

**1** ARC Centre of Excellence in Plant Cell Walls, School of BioSciences, University of Melbourne, Parkville, Victoria, Australia, **2** Bioinformatics, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia, **3** School of Mathematics and Statistics, University of Melbourne, Parkville, Victoria, Australia, **4** School of BioSciences, University of Melbourne, Parkville, Victoria, Australia, **5** Bioinformatics, Murdoch Childrens Research Institute, Parkville, Victoria, Australia, **6** Life Science Computation Centre, Victorian Life Sciences Computation Initiative, Carlton, Victoria, Australia

* andrew.lonsdale@lonsbio.com.au





**Citation:** Lonsdale A, Sietsma Penington J, Rice T, Walker M, Dashnow H (2016) Ten Simple Rules for a Bioinformatics Journal Club. PLoS Comput Biol 12(1): e1004526. doi:10.1371/journal.pcbi.1004526

**Editor:** Fran Lewitter, Whitehead Institute, UNITED STATES

**Published:** January 28, 2016



**Funding:** The authors received no specific funding for writing this article.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

As science becomes increasingly interdisciplinary, we are expected to acquire both breadth of knowledge and depth of expertise. In bioinformatics, this is especially true. Keeping up to date with major techniques across multiple specialisations can be daunting, but you need not face it alone.

A journal club is an excellent way to take in the scientific literature, keep up with developments in your field, and hone your communication and analytical skills.

In general, a journal club is a group of people who meet regularly to discuss one or more scientific papers. The structure of such a club can vary. In the more traditional format, an individual studies a paper and then presents it to the group, usually in the form of PowerPoint slides, with time for questions. In some institutions, the journal club is for students only, designed to fulfill the requirements of a course or postgraduate program; attendance is obligatory, the scope of the literature is narrow, and the format is prescribed. The preparation of slides and a lecture may be required. Other kinds of journal clubs are just lab meetings in disguise, with the usual lab head and group members in attendance and one member nominated to present the paper.

A formal style often fits well within an established academic structure—but what if your discipline is emerging?

Consider the field of bioinformatics. Expertise may be spread across departments and institutions, and there may not be an obvious place or critical mass in any one lab for a traditional journal club. Research students, "pet bioinformaticians," [1] and those interested in bioinformatics from adjoining fields all need a place to gather.

We are pleased to offer an alternative structure to address this situation—an informal journal club, designed to bring together a diversity of backgrounds and career stages to discuss bioinformatics while building a network of like-minded peers. Additional benefits of such a journal club may include friendship and breakfast (see Rule 2)! We thoroughly recommend it to anyone who asks (as well as those who don't).

While this advice is drawn from our experiences in the Parkville Bioinformatics Journal Club, it is applicable to developing informal journal clubs of all disciplines. The advice contained in these rules will also help those who want to spice up their existing formal format.

So don't be a "lonely bioinformatician"[1], create a journal club! Follow these Ten Simple Rules to find out how.

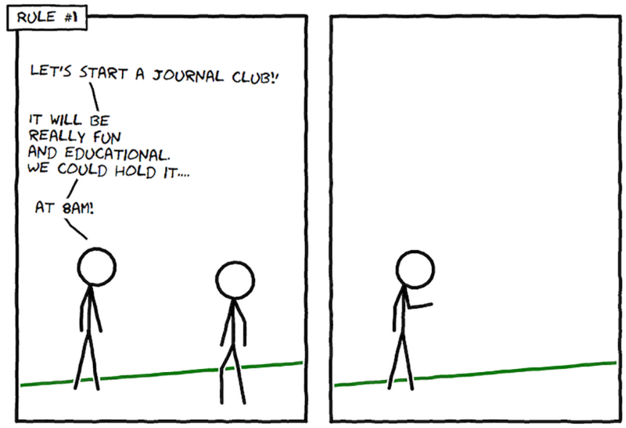



**Fig 1. Have it at 8 A.M.! This comic was drawn using Comix I/O (http://cmx.io/).**

doi:10.1371/journal.pcbi.1004526.g001

## Rule 1: Have It at 8 a.m.

Our research has indicated that the best time to hold a journal club is at 8 A.M. ($n = 1$). The most common reaction to this conclusion is best summarised in Fig 1. Words like "lunchtime," "evening," and "weekend" are likely to elicit similar reactions and, along with 8 A.M., seem unlikely times for a journal club. However, there is a method to our madness.

Forming an interdisciplinary journal club typically requires volunteers from various areas with competing schedules and commitments. This makes finding a suitable time challenging. We recommend you find a time that is typically free for the majority and suits both students and early career researchers, who may have time commitments during "normal" working hours. For us, that time was 8 A.M.

We chose our meeting time by polling members of our bioinformatics community. Any club consists primarily of its members and requires a culture of willing participation to run well. Mere attendance will not do the job. A time outside regular commitments may appear difficult at first, but attendance indicates that members truly wish to participate; otherwise, they would not make the effort to appear at all!

The caveat to this is that this noncompulsory attendance results in only people of a certain disposition attending; in this case, those for whom an 8 A.M. optional journal club is appealing. This is a self-selection bias that we can live with.

An early start means that members are neither tired from nor distracted by the regular issues and chores of the day. There was also the bonus that it meant that members were on campus early in the morning at least once a fortnight—a vast improvement for some of the authors.

If those interested in starting a journal club are early risers, then follow our lead, but be sure to meet the needs of your own community. In setting up an unconventional journal club, there is scope for unconventional structure.

## Rule 2: Keep It Casual (with Coffee!)

Having decided when you will hold your journal club, a location is the next requirement. For an 8 A.M. start (Rule 1), the priority for the caffeine-fueled organisers was finding a venue with great coffee. This initial choice of venue helped form additional requirements. Adding a journal club to already busy schedules can require an incentive. For our members, breakfast and coffee made the early start bearable, so a venue with a variety of food and drink, including coffee, was essential. Available tables (without the overhead of booking) and proximity to the campus

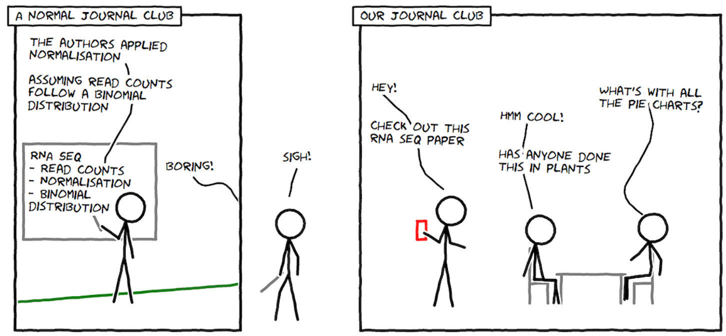



**Fig 2. Why our journal club beats a "normal" journal club.** This comic was drawn using Comix I/O (http://cmx.io/).

doi:10.1371/journal.pcbi.1004526.g002

precinct were also desirable traits. This placed the club in a cafe, which also ruled out formal presentations with PowerPoint slides (much to our relief), and this choice of location shaped our journal club (Fig 2).

The choice of an 8 A.M. start helped form the mood and style of our journal club, and the time that suits your community will do the same.

If you are holding your journal club in a meeting room, consider some other kind of incentive to make your club more appealing and encourage people along. For example, a rotation of baking or bringing along snacks. If it is at lunchtime, include lunch as part of the schedule. Since the club is (ideally) relaxed and social, it relies on its members being genuinely enthusiastic in order to function, so a style that suits your community is essential.

Over the course of journal club meetings, our style quickly developed. An informal location with food or drink lends a welcoming air. The early start and ability to order food and drink also meant that 8.10 A.M. became the de facto start time, as members tended to arrive on time and then place an order; this set the relaxed tone that followed. The leader of the session gives a short overview of the paper, explains what drew them to choose it, and then opens it up for discussion. That is as much structure as we needed.

This style makes it easy for new people to attend. They can just sit and listen until they feel comfortable enough to contribute. Latecomers are not admonished, and on more than one occasion, participants admitted to having not read all (or any) of the selected papers! This kind of admission is unlikely to be well received in a formal club but is met only with gentle teasing in ours.

Where you decide to hold your journal club will affect the style and format, so ensure your venue matches the type of club you are aiming for.

## Rule 3: Multidisciplinary Is Not a Dirty Word

As we've alluded, the traditional journal club format often isn't a good fit for bioinformatics. Our members come from various universities and institutes, many from groups in which they are the "lone bioinformaticians" or where all their lab mates are computational and they are just learning to program. They feel alone in their own labs, so they come to a journal club to feel part of a larger field.

This group may call themselves many different things (e.g., bioinformaticians, computational biologists, systems biologists) and have different areas of expertise. We have molecular modelers, computer scientists, mathematicians, and cancer genomics gurus. Add to this mix biologists seeking more bioinformatics knowledge, and the range of expertise is even broader.

This multidisciplinary context is a challenge, but it is this diversity that is part of what makes our journal club so special.

It can be intimidating to come into a group where everyone knows each other or everyone is expert on the same topic and you are not. The breadth of expertise is an advantage in this case; our diverse group means that for every paper, there will be someone who knows nothing about the topic and others who are near-experts.

By having an unconventional approach and an informal style, we create a safe environment where everyone feels comfortable asking questions when they don't understand, and others flex their teaching skills by answering these questions. With so many people coming to new fields of study with their own unique backgrounds, we inevitably discuss harebrained solutions to seemingly insurmountable problems. We encourage questions, opinions, and tangents. Ensure the group is diverse by inviting people from a range of labs and departments. We have an open membership approach and encourage attendance from different cohorts. Mixing research students, coursework students, and early career researchers helps give perspectives from differing stages of career and life!

A mixed group can prevent situations where complex methods might be skipped because they are outside the expertise of the group, as what may appear to be complex to newcomers can often be explained succinctly by those familiar with it.

By having a diverse group of people in which someone is almost always bound to know an answer, you create an environment that is valuable to someone trying to learn about other disciplines. Many of us involved in bioinformatics were at one stage computer scientists without a biology background or biologists without computing or statistics. Expertise in everything is unlikely, and there will inevitably be gaps.

The diversity of a journal club is a way to fill in those gaps and benefit from our shared knowledge.

## Rule 4: Let Your Topics Be As Diverse As Your Members

With a regular time and venue arranged, the topics to be covered in your journal club need to be defined.

Though some journal club formats have a narrow scope of topics, restricted to the research focus of a lab or from a prescribed list, there are benefits to having a more relaxed approach. Given the diverse backgrounds of our members (see Rule 3), and our informal approach, we do not prescribe topics.

Members are free to volunteer to lead on a topic of their choosing. This flexibility allows for a range of discussion styles—teaching from experience in one's own field, branching out into something new to everyone, or even using the journal club as a crowd-sourcing opportunity for a critical paper in your literature review.

This flexibility has led to a broad range of topics. Scientific topics have included synthetic biology, glycoinformatics, ChIP-seq, the human microbiome, systems biology, CRISPR-Cas9, and molecular modelling.

A variety of topics can also act as good entry points to your club. A constant schedule of scientific topics may intimidate new entrants to the field or those who are curious about your club. The occasional topic that focuses on general skills for science, or issues common to all disciplines, can encourage new members and reinforce that your club is atypical.

We have discussed overcoming procrastination, writing more effectively, online learning, editing Wikipedia, student groups, Software Carpentry, document preparation systems for science, and even the PLOS Ten Simple Rules collection itself!

Leaders are free to choose areas of research that they know well and will enjoy dissecting but need to remember that the rest of the group may not share their expertise or personal interest. It is often helpful to accompany domain-specific articles with a suitable review paper to introduce others to the important concepts.

Frequently study papers that are neither fashionable nor trending. The major breakthroughs of tomorrow are among the peripheral topics of today, and the mainstream is, almost by definition, slow to recognise the truly revolutionary. This was recognised by Baez [2]: "Look for problems, not within disciplines, but in the gaps between existing disciplines. The division of knowledge into disciplines is somewhat arbitrary, and people put most of their energy into questions that lie squarely within disciplines, so it shouldn't be surprising that many interesting things are lurking in the gaps, waiting to be discovered."(Aside: this blog also made for a good journal club meeting.)

Even papers destined for obscurity provide a new perspective, and some of our most stimulating discussions were based on papers from adjoining fields such as genetics and information theory.

## Rule 5: Find Good Articles For Discussion

A high-quality journal article doesn't necessarily make for a good journal club topic. It's important to read good science, but journal club is also about honing your analytical skills. Find papers you respect enough to criticise. Look for papers with interesting ideas, issues, or challenges that you can discuss.

The articles you choose should be specific enough to gain interest, yet flexible enough to engage the diverse members of the group. It is good to choose a challenging article, but not too challenging. If the paper is dense, it should also be short. Remember that your members are most likely reading this article in their spare time.

For the same reasons, avoid big papers: it is hard to digest and critique ten years' worth of work over one coffee (or even two).

How do you find topical journal articles? RSS feeds can be used to suggest relevant articles, as can reviews. When choosing articles, accessibility can be an issue, particularly if you have members who don't have access to big university libraries. Consider choosing articles from open access journals.

In an interdisciplinary club, members will have various backgrounds. Consider additional resources for varying levels of expertise. A selection of the current literature could optionally be accompanied by links to seminal papers or online primers such as Wikipedia to give context.

Papers don't have to be on paper. Blogs, social media, and other online resources can add variety to your discussion diet. Twitter and blogs can be particularly useful if you're looking for something topical or contentious (Fig 3). Some blogs have a certain renown for applying strong criticism to published journal articles. It can make for lively conversation to first read an article that seems like valid work from leading people in a field, and then read a blog post that tears the same article to shreds.

Social media, post-publication peer review, preprints, and open access are changing the way we communicate scholarly information. Let them change the way we select articles for journal clubs as well.

## Rule 6: Make Leading Easy

As people join your journal club, expanding the roster of leaders is important. An informal club relies on volunteers, so making it easy to lead a session is critical. Founders of the club

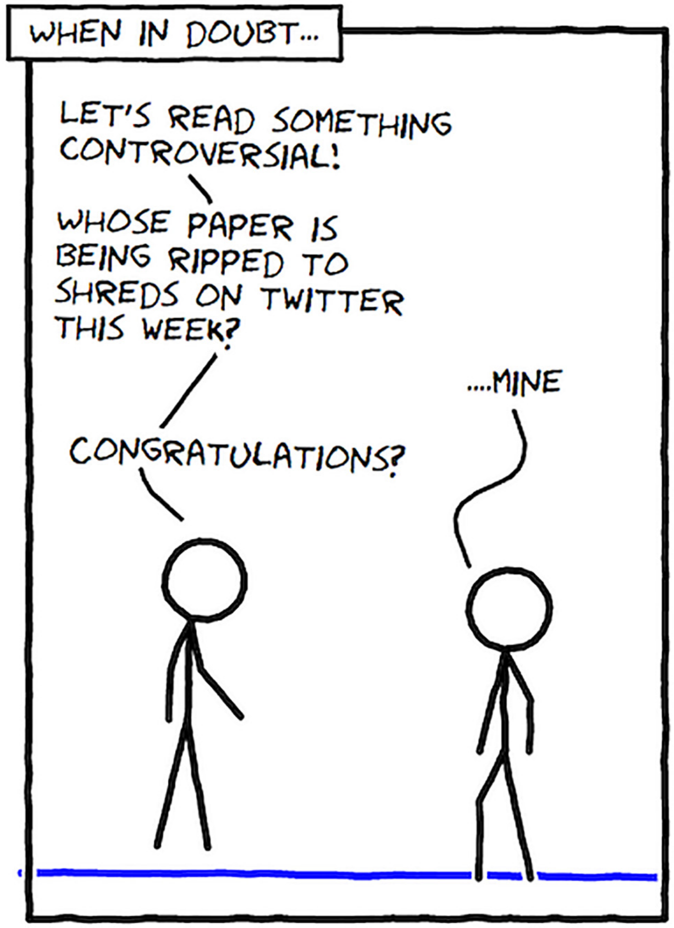



**Fig 3. How to find contentious articles.** This comic was drawn using Comix I/O (http://cmx.io/).

doi:10.1371/journal.pcbi.1004526.g003

may lead more often than others, but a club works best when everyone shares the article selection and leadership.

Moving from participating to leading can be daunting. It need not be. We keep the format simple and provide a quick-start guide to get new members up to speed. A short document or a blog post (such as this one we prepared earlier: http://parkville-bioinformatics-journal-club.blogspot.com.au/2014/01/leading-journal-club_28.html) with advice on how your club runs is a simple way to impart advice. You can also use this document to elaborate on the benefits of leading a journal club session.

Each journal club leader brings their own personality to the task. One memorable opening line, "Welcome to Journal Club," remains the best and only formal start to a journal club so far, but it all depends on the style. In general, introduce yourself, and ask for a brief round of introductions. A simple introduction at the start can make newcomers more relaxed about joining in, and help those with failing memory remember names.

One benefit of leading journal club is you can use it to forward your other academic goals. The discussion at journal club could provide motivation and fresh perspectives for future projects, and any material prepared for the discsussion can be fodder for your next literature review, journal article, or blog post.

## Rule 7: Avoid Dead Air

Like good radio, a journal club works best when there is constant sound. Newcomers should not feel pressured to participate prematurely, but try to keep the conversation going when those inevitable lulls occur. This is primarily the leader's responsibility, and it helps to bring a list of questions or comments.

Give a brief summary of what the paper was about. Think of this as a verbal abstract, but with even less detail. The idea is to give people who haven't read it (or have only skimmed it) a basic idea of what the paper was about. You may want to finish with a comment about what you thought of the paper. What were its strengths and weaknesses? Why did you choose it?

However, the leader is not alone. All members should feel able to chip in when there is a gap, and like commercial radio, there is nothing wrong with the odd cliché, e.g., "What are the next steps from this paper?", or as one of the authors (AL) asks almost every week, "Has this been done in plants?"

Avoiding an awkward silence need not be a chore. In an informal journal club, going off topic is encouraged! Feeling comfortable enough to ask questions that you would not ask in a more formal setting is a sign that your journal club has the right spirit.

## Rule 8: Be Organised

A journal club of only one person is referred to as "reading alone." Success depends on people who turn up consistently and volunteer to select papers, so it helps to stay organised.

Communicate regularly with email, Facebook, or cans and bits of string—whatever works, or use them all! What is the best way to distribute papers? Google Drive, or email PDFs? Emailing links has sometimes been fraught with problems. Sorry, we don't know the optimal answer for this one!

Ongoing recruitment and encouragement is important—otherwise, journal club will become a shrinking clique. Advertisements, Facebook, and reminder emails all have a role, and remember to make it easy for new leaders (see Rule 6).

Informal does not mean improvised. Schedule leaders in advance and announce articles to give time for everyone to read them. Always know when the next meeting is and who is responsible for choosing a paper. For example, our club found that after the end of year break, holding a meeting at least symbolically in January (summer break in the southern hemisphere) ensured that the club was able to continue momentum into the new year.

Keeping track of the next meeting means you need to keep records. When are the meetings? Who led, how many attended, and which papers were discussed? This is useful not only for fancy graphs (see Fig 4) but also for letting newcomers know about the history of the club and which topics have already been covered. It also means that you can check your facts and find citations if you need to write a Ten Simple Rules for a Bioinformatics Journal Club.

## Rule 9: Be Adaptable

One particular strength of an informal journal club is its ability to adapt to the needs of its members.

For example, when scheduling your journal club, be sure to consider all of the factors that impact your members, such as holidays, exams, PhD confirmations, and grant writing. If necessary, reduce the frequency of meetings at such times.

It's easiest to organise around weekly meetings but hard to arrange a different leader and read so many papers every week. If, like us, you find a fortnightly meeting easier to sustain, then it is useful to have something else in the intermediate weeks. Our club has a writing gym ("pomodoro" sessions) modelled after Shut up and Write! (http://thesiswhisperer.com/shut-

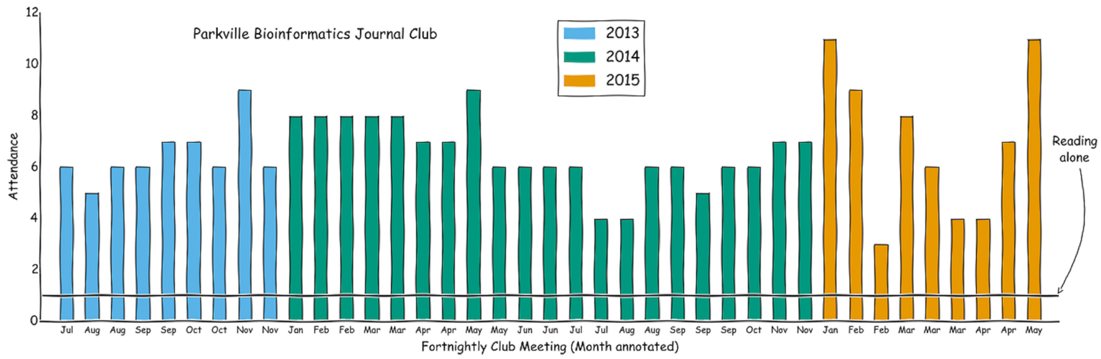



**Fig 4. Journal club attendance over time versus reading alone.**

doi:10.1371/journal.pcbi.1004526.g004

up-and-write/) on the alternate weeks, but yours might have different needs. Having that timeslot reserved for your group every week will help people trust that they can always show up at that time and catch up with colleagues over a coffee.

Although we didn't originally envisage journal club as a platform for collaborations, we've found that meeting regularly to write amongst a group of supportive peers has led to a publication, with a style that suits the informal atmosphere of the club [3].

Viewing the club as more than just a journal club has been an invaluable chance for exposure as early career researchers. An earlier version of these rules appeared as a poster at the inaugural Australian Bioinformatics Conference (http://bioinformatics.net.au/abic2014/) and is available at F1000Research (http://f1000.com/posters/browse/summary/1097136).

It turns out that journal club can be more than just reading papers, which leads us to Rule 10.

## Rule 10: It's about More than Just the Articles

The encouragement to read and the opportunity to discuss are why we meet, but the biggest gain is actually a community of peers.

The most exciting bioinformatics journal clubs consist of members with diverse backgrounds, acting as a microcosm for the field of bioinformatics at large. Our journal club members come from all over the field of bioinformatics.

A diverse background of journal club members makes for a diverse selection of literature, acting as revision for some and revelation for others. The articles we discuss not only improve our understanding of bioinformatics but also broaden our vision of what bioinformatics contains.

Drawing together a heterogeneous group of students and early career researchers on a regular basis creates a network that spans the existing structures of labs, departments, faculties, and schools. A vibrant journal club becomes something more than the sum of its parts. It becomes a community.

Coffee, articles, friendship, news, research, support, and ideas merge. You can measure the success of a journal club by more than just how many attend, or the impact factor of the articles selected. Your journal club will be successful when you don't even think of it as a club at all.

## Conclusions

Remember to enjoy yourself. In an unconventional journal club, you are not giving a formal presentation. No one is judging you, but if they are, give them a copy of these Ten Simple Rules.

Create an informal atmosphere for your journal club. We do this using coffee, but you need to find what works for your community. You may wish to complement your journal club with other activities such as a writing club or social events.

Choose a venue that matches the intended style of your club. Invite and encourage a diversity of backgrounds and expertise to your journal club, and let the topics and articles reflect that diversity.

Keep the barrier to entry low. Because attendance is not compulsory, journal clubs like this rely on volunteers to choose the papers and lead the session. It needs to be easy for people to do this. There should be simple and easy-to-follow guidelines to help members move from observers to active participants (see Rule 6).

Reading bad papers can be good. Whereas your supervisor may tend to focus on fairly specific and conventional literature, a journal club is a chance to discuss articles that may be controversial or even flawed. Don't be afraid to look outside your field for a fresh perspective. We also recommend supplementing journal articles with less formal commentary such as blogs.

Finally, we would like to leave you with this thought: journal club is not really about reading and discussing journal articles. Sure, that's what we drag ourselves out of bed to do at 8 A.M. on a Tuesday morning, but the true purpose of journal club is to build a bioinformatics community that spans institutes, ages, and even disciplines.

Journal club is about more than just the articles; it's about the people you share them with.

## Acknowledgments

Some of the ideas used in this article first appeared on the Parkville Bioinformatics Journal Club blog: http://parkville-bioinformatics-journal-club.blogspot.com.au/. The authors would like to thank the Dr Dax Kitchen for their coffee and great service (putting up with us).

We would like to acknowledge the enthusiasm of everyone who has volunteered to lead in the Parkville Bioinformatics Journal Club: Andrew Lonsdale, Harriet Dashnow, James Harrison, Jane Hawkey, Min Wang, Jocelyn Sietsma Penington, Lavinia Gordon, Michael Walker, Tim Rice, Marek Cmero, Lesley Raven, Scott Ritchie, Artika Nath, Thomas Coudrat, and Alexis Lucattini.

We would also like to thank Mina Stevenson and Alicia Oshlack for their feedback on the manuscript.

## References

1. M Watson. A guide for the lonely bioinformatician. 4 April 2013. In: opiniomics. https://biomickwatson.wordpress.com/2013/04/23/a-guide-for-the-lonely-bioinformatician/
2. J Baez. Finding and Solving Problems. 18th February 2014. In: Azimuth Project. http://johncarlosbaez.wordpress.com/2014/02/18/finding-and-solving-problems/
3. Dashnow H, Lonsdale A, and Bourne PE. Ten simple rules for writing a PLOS ten simple rules article. *PLoS Comput Biol*, 10:e1003858, 2014. doi: 10.1371/journal.pcbi.1003858 PMID: 25340653

PLOS | COMPUTATIONAL BIOLOGY

Editorial

# Ten Simple Rules for Starting a Regional Student Group

**Avinash Kumar Shanmugam[1], Geoff Macintyre[2], Magali Michaut[3], Thomas Abeel[4,5]***

**1** Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, **2** NICTA Victoria Research Laboratory, University of Melbourne, Melbourne, Victoria, Australia, **3** Computational Cancer Biology, Netherlands Cancer Institute, Amsterdam, The Netherlands, **4** VIB Department of Plant Systems Biology, Ghent University, Ghent, Belgium, **5** Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

Student organizations are a great way to network and take a break from the rigors of the classroom. They provide a range of benefits beyond regular coursework and can be critical to having a well-rounded education. Many students are active in organizations at an undergraduate level, but the increased demands of a master's or PhD typically result in reduced participation at a graduate level. However, a student organization can equally provide benefits for a graduate student, especially if it is centered on the student's area of study. In this article, we focus on Regional Student Groups (RSGs). An RSG is a group of like-minded students across a geographical region with a common field of research. The group provides a support network and collaboration opportunities via a collection of individuals who "speak the same language." The RSG concept was created by the International Society for Computational Biology Student Council to address the needs of students in the field of computational biology in each region. Currently, the RSG program consists of over 20 regional student groups worldwide. In this article, we provide ten simple rules for how to start a regional student group in the hope that others will start up similar groups around the world.

## Background

The International Society for Computational Biology (ISCB) Student Council (SC) is an organization dedicated to nurturing and assisting the next generation of computational biologists. The SC offers networking opportunities and soft skill training to scientists in bioinformatics who are in the early stages of their careers. This is achieved through a number of activities, including the long-running symposium series organized in conjunction with the Annual International Conference on Intelligent Sytems for Molecular Biology (ISMB) and the European Conference on Computational Biology (ECCB). While successful in uniting students across the globe, the international symposiums did not provide opportunities that directly addressed the needs of computational biologists at a local level. Thus, a RSG program was established.

Since 2005, the RSG program of the ISCB Student Council has provided the opportunity for local student groups to become affiliated with the SC to share experiences and resources (http://www.iscbsc.org/content/regional-student-groups). RSGs typically represent the student community of one country, although we also have examples of supranational and subnational RSGs. Over the course of the last seven years, they have continued to grow and flourish. With over 2,000 students across 23 countries, the RSGs are providing valuable initiatives to support and promote students in bioinformatics and computational biology at the local level.

RSGs fill the gap between the global ISCB and SC organizations and the institutional associations that exist at many universities. Computational biology is a cross-disciplinary field, and young researchers in bioinformatics are often spread across the country in different groups with sometimes only a few bioinformaticians per institute. Having a country- or statewide association provides the critical mass to organize meetings, share ideas, and interact with peers. RSGs have a big advantage compared to a global society in that they can leverage local strengths and needs to offer a tailored program, yielding high impact.

The following ten rules are guidelines on how to get started with your own local group of trainees to form a successful local bioinformatics association. These rules are derived from the SC's experience with setting up and guiding RSGs in 23 different countries over several years. The rules are kept as generic as possible so they can be beneficial to other multidisciplinary areas as well.

## Rule 1: Form the Right Team

Finding the right people to start your group is crucial for its success and survival. As with all volunteer organizations, the workload needs to be shared among individuals, and initially, work will need to be put in without immediate reward. Everyone involved must *want* the organization to succeed and must be willing to invest time and effort. Ideally, this group of people should include one or more persons who can take initiative and set things rolling; good communicators who can get other people excited about the group and get them involved; a visionary who sees where the group should go; and finally, people who can take responsibility for dotting the i's and crossing the t's. The right set of motivated people will see the group through the first year, and the eventual rewards will be worth the effort.

It is a good idea to have a formal leadership committee that will take responsibility for activities. A collective of hard workers allows for a dynamic and nimble team. This group of people should be as diverse as possible in terms of experiences, ideas, career stages, etc. In the beginning, it may be easier to plan with a small group of friends to get the project started, but it is essential to broaden your horizons as soon as possible to increase the chances of success. In addition, it is helpful to have at least one faculty advisor or senior mentor who

**Citation:** Shanmugam AK, Macintyre G, Michaut M, Abeel T (2013) Ten Simple Rules for Starting a Regional Student Group. PLoS Comput Biol 9(11): e1003340. doi:10.1371/journal.pcbi.1003340

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** November 21, 2013



**Funding:** The authors received no specific funding for this article.

**Competing Interests:** I have read the journal's policy and have the following conflicts. Thomas Abeel was a member of the ISCB Board of Directors from January 2011 to January 2013. Avinash K. Shanmugan is a member of the ISCB Board of Directors from January 2013.

* E-mail: tabeel@broadinstitute.org

serves in an advisory role. They will help with getting access to resources and people, and having a senior advisor provides your fledgling group with increased credibility and legitimacy.

Besides the leadership committee, which would typically consist of three to four people, there should be a larger core committee. Core committee members should be located around the country or region (different cities, different universities) to make the organization truly representative of the region. Again, this will help with getting access to various resources and provide inside contacts for advertising later on.

Be careful to find the right people. In the end, putting in the time and effort to find enthusiastic and motivated people will be well worth it.

## Rule 2: Have a Plan

The basic goal of all regional student groups is to create a student network and work towards providing activities that benefit members. To achieve this, it is important to identify the needs of the community. For example, what do potential members want: better industry contacts, better information sharing, more meetings, or a support network of peers? The community of interest can be engaged on the sidelines of conferences or other talks and events in your field. If there aren't any such opportunities to meet with the community in your region, you can solicit community feedback by e-mailing flyers to universities to put up on noticeboards. Interacting with the community can be as simple as chatting with your labmates or picking up the phone and talking to colleagues. It is all about getting out from behind your computer and talking to other people. Creating opportunities for community interaction should be on top of your to-do list.

Once these needs have been identified, a plan should be put in place to attempt to fulfill those needs. The plan should be thorough and concise—don't write a novel, but ensure that sufficient details are given so that the plan can be put into action. Following are three aspects that should be covered in the plan.

1) The practical plan: What concrete things does the group plan to do in the next year? These need to be outlined with clear deadlines and measurable progress milestones. An easy way to do this is to identify the end goals to be completed in the year and then work backwards from these deadlines to fill in the milestones needed to get there. Make sure all goals are obtainable.

2) The aspirational plan: In general, you want to answer a few strategic questions: What does the organization want to achieve in five to ten years? How will it impact the community? How is the group working towards these goals? This should be more about setting a philosophical tone of what the RSG should achieve and act as a wish list. Some of the goals you want may not be possible right now, but do not restrain yourself. This is where you can dream about the future.

3) The succession plan: *The king is dead. Long live the king!* Finding a replacement leadership team is one of the biggest hurdles faced by regional student groups. Students may graduate or relocate, or may just become otherwise busy with their academic lives. To build a robust student organization, you need to plan to deal with some level of volunteer turnover. This is especially critical for the leadership roles. In the past, some RSGs have gone dormant after the team members were unable to meet their obligations to the organization and did not plan for their succession. The next generation of organizers has to be recruited while the current team is still active, so thinking about a succession plan early is a good idea. At every event, devise a plan to recruit new members and give them new responsibilities. Eventually, some of these members will become part of the next leadership committee. Creating a core committee (see Rule 1) is also very useful in this respect because it allows you to have a pool of people that are essentially leaders in training, lets people grow into their role, and helps maintain continuity.

## Rule 3: Organize Events

Events are a great way to engage the community, increase visibility, and get new members. Working towards an event can really help bring your leadership team together. A good rule of thumb is to use a big event to gain high visibility and follow up with a number of smaller events to ensure people keep coming back; i.e. consolidate your community. The large event may be an annual symposium or workshop. The smaller events could be as simple as a bi-monthly journal club that rotates through various universities in your area or a social event over a meal with a beer or a glass of wine. What is better than yeast to make people talk and like each other?

You have to be creative in the types of events you organize, and there is no rule set in stone for what will work for your local demographic. The type of events that will work are very dependent on your local needs, so get input from the community and plan accordingly. Soft skill training—such as the art of presentation or techniques for scientific writing—is generally well-received because it is lacking in almost all curricula around the world, so workshops for those skills tend to draw a large audience. Programming-related workshops and/or competitions are quite popular in many regions, but there has been equal success with breakfasts, essay writing, or quizzes. Remember that many students participate in an event to gain experience that will give them an edge when they apply for a job. Consider running events that help students in the quest for a job in academia or industry. Be creative, listen to your members, and you will deliver some great events.

## Rule 4: Think Frugal

For a student group, raising funds for activities is always a challenge. Establishing relationships with institutes, companies, or national granting agencies creates opportunities for acquiring funding or support. Getting cold hard cash can be difficult. However, it can be easier to obtain support in kind: for example, using an auditorium or classroom for free, or getting free products from a company to use as prizes. Some speakers might be willing to cover their own travel for speaking at a student event, or a university may cover the expenses of a speaker if all university employees can attend. Seek out and utilize opportunities to get the most "bang for your buck." For instance, partner your annual event with a popular conference in your region or country. This will help you save on venue costs and will guarantee an audience.

## Rule 5: Think about Resources and Logistics

Once the core team has been formed, the next step is turning the organization into a recognizable brand. Just like with a sports team, the team needs a name, a logo, a web presence, and mailing lists. The list goes on. Just remember it is not all about brainstorming crazy ideas for a name or having fun designing a cool logo. It is about ensuring successful communication. Many practicalities have to be considered when setting up modes of communication, but you can be creative in finding the required resources. Where do you put the website? At the home institute of one of the committee members? At a free commercial service like Google Groups or with a paid service? Weigh the alternatives. Along with a website, it is recommended to set up an online identity at various social media sites including LinkedIn, Twitter, and Facebook. Make sure that the entire leadership has access to these to update them.

For day-to-day operations, most planning happens online, so there is limited need for physical logistics like meeting rooms. Most RSGs use Skype or other conference call technologies for meetings.

Resources for events, however, are a completely different beast. Events generally need their own name, logo, and web page. The hardest part is getting the physical resources sorted. You may need to take care of meeting or conference rooms, a submission system, poster boards, A/V equipment, etc. (see Rule 4). The location of your event is going to be a big factor in its success. The venue should be readily accessible by your target audience. If the majority of people attending your meeting arrive by car, make sure there is plenty of (free) parking available; if they will come by public transport, make sure your venue has good connections and that service runs before and after the scheduled times of the event. You do not want to kill an extremely successful event with a two-hour walk home from a remote place at night in the rain…

Involving people who are local to the event venue often makes your life easier. They are your boots on the ground to get everything moving smoothly. They can walk up to anyone in person at the venue and may immediately know the right person to get any issues sorted out.

Last but not least, we are all human: if your event lasts more than an hour, you need to think about basic needs like food, drinks, and restrooms. Contrary to popular belief, scientists need more than just caffeine to get through the day.

Organizing a conference requires significant investment in terms of time and planning. It may be easier to start with something smaller that does not involve so much up-front work to get the ball rolling. A practical training workshop, a journal club, a social gathering, a quiz…the possibilities are endless.

## Rule 6: Promote the Organization

When getting started, create a flyer about your organization that you can send to universities with a bioinformatics program and ask for it to be put up on their notice boards. If you do not have funds to send hard copies by post, you can just e-mail a PDF copy and ask for it to be printed out and put up. This is easiest if you have a core committee member at the target university who can print it and put it up (see Rule 1) or, even better, present it themselves to the new students during a course. It may be useful to keep an up-to-date list of people responsible for master's or graduate programs that you can contact every year to remind them that your group exists and ask them to promote it to the new students.

Approach the organizers of popular conferences or meetings in your region to get a time slot to give a (short) presentation about your RSG during the meeting: for example, during the opening remarks or before or after a coffee break. Even a slide about your organization in the cycle of slides that are shown in the breaks between talks may be useful. Approach local science magazines or the science sections of newspapers about whether they could run an article about your RSG. You can also put announcements in institutional newsletters or magazines. However, make sure you have a single point of contact where people can find information about the organization, about future events, where to sign up for regular news, etc. (see Rule 5).

The best way to promote the society is by word of mouth. Get out there and talk to fellow students at conferences, workshops, meetings, etc. Not all students will remember to follow up. Make sure you get their contact information and follow up after the meeting. Consider creating a newsletter or just a periodic e-mail alert to help members keep up to date with what is happening. It is also useful to update nonmembers to give them an incentive to join. Until your RSG attains a critical mass of people where it can become somewhat self-sustaining, you will need to work to push this information out to the community and actively recruit people to join.

Once people are coming to your events, you need incentives to keep them coming. There are two main incentives that are easy to have in any event, regardless of topic or theme: CV lines and networking. First, paper and/or poster awards, travel fellowships, etc. are things that are easy to provide, but give distinct benefit to the winner(s) in that they can include these on their CV to further their career. Second, maintaining a good social component to your events will make sure that people attend even when the event does not have a theme that is directly related to their research.

## Rule 7: Document Decisions, Methods, and Events

This sounds very serious…and it is! Because you want your group to grow and live a long time, you need some serious documentation. A student group should expect to see a fair amount of turnover. As mentioned earlier, students get *real* jobs, get busy, or just move on to other things. Therefore, it is very important to document discussions, decisions, and operating methods to help maintain continuity. The simplest way to maintain records is to conduct discussions by e-mail or any other electronic medium that has automatic archives. It is useful to use a mailing list with all of the leadership team for discussions rather than using personal e-mails. Not only does this make it easier to look up information later, but it also promotes transparency. For any offline discussions, take down quick minutes and share them on the mailing list. As your group matures, you can move to records that are more systematic.

After a larger event, it may be worthwhile to report on it. On the one hand, this recognizes the people who put a lot of time and effort in the event, and on the other hand, it also helps to promote your group (see Rule 6). Various journals offer a means to publish meeting reports, and local newspapers are often happy to run a story on local events. Having regular reports on your events will help with recruiting more members and will assist in gaining support to further expand (see Rule 8).

## Rule 8: Find Support and Benefactors

Once you have the basics in place, you can start looking towards the future. Aim to create relationships with other organizations, be they companies, universities, or other professional societies, because they can help provide funding and/or material support. Being associated with an established organization can help you apply for grants to support your work, too. For example, the SC has a funding program for RSGs where up to five projects are funded each year. Finding monetary support is going to be challenging, and you will need to be frugal (see Rule 4) with whatever funds you get together.

Make sure to cultivate good relations with local professors and administrators; they are your keys to accessing resources: using rooms at the University and convincing renowned scientists to speak at your event, for example. These personal connections will help you tremendously with acquiring logistical support for your group. Getting connections to companies while fundraising may even prove to be beneficial when looking for a job later on.

## Rule 9: Be Inclusive and Reward Contributions

One important step to make sure your RSG is healthy is to involve as many people as possible in actively contributing roles, not just as participants in your events. You should think how to motivate and incentivize people. Find ways to be inclusive and integrate people in the organization.

Many people will start participating out of interest and self-motivation, but competing priorities can get in the way. You must think of ways to keep people involved in the long run. Proper recognition for effort and involvement will make sure people want to volunteer to do a job again. Make sure you thank everyone by name in front of the event participants—bonus points if you can mention some of their actual contributions. Make sure that key contributors to your event have a title, e.g. "program chair," for the person putting the agenda together and organizing the abstracts. This ensures their work can be added as a line on their CV and enables them to justify the time spent on doing this job.

Allow people to be involved no matter how much time they can contribute. Not everyone will have the time to be involved in the RSG at a leadership level. Try to find avenues where they can still contribute to the organization with only a minor commitment. Some examples of low-commitment activities may include reviewing abstracts for a symposium, answering questions on the mailing list, and forwarding useful information to the mailing list. Sometimes these tasks can be done more efficiently by a few dedicated people, but splitting it up and letting more people contribute can help create a feeling of belonging and ownership within the community. Even though involving many people is not the most efficient in the short term, it will likely improve the success of the RSG in the long term. Some of these minor contributors could get motivated to take on more significant positions later on.

## Rule 10: Less Talking and More Doing

You do not want to get lost in translation. Do not spend all your time making grand schemes to start the ultimate RSG or organize the most spectacular symposium ever. It is extremely important to get started. While you should not rush in headfirst, you also should not spend too much time planning every single aspect and contingency in minute detail. You may plan forever and get nothing done. This fine balance is easily missed, and it happens all too frequently that groups are trapped in endless discussions without ever moving beyond talk. Just think of something small you can do and get it done. It is better to do something that you can later improve than to envision something amazing that will just never happen. *Release fast, iterate often.*

## Conclusion

In addition to the considerable benefits that an RSG can provide to a student community, the authors can also attest to the fact that being involved in setting up an RSG will be an educational, transformative, and rewarding experience.

We hope that this short guide on starting up and running regional student groups will help inspire young scientists to band together and build the next generation of computational biologists.

Editorial

# Ten Simple Rules for Providing a Scientific Web Resource

**Sebastian J. Schultheiss***

Machine Learning in Biology Research Group, Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany

Many projects in computational biology lead to the creation of a small application program or collection of scripts that can be of use to other scientists. A natural progression is to make this tool available via a Web site or by creating a service for it, from now on collectively called "Web resource."

We conducted a survey among providers and users of scientific Web resources, as well as a study on availability. The following rules reflect the experiences and opinions of over 250 scientists who have answered our questions and who use Web resources regularly, as well as our own experience. The study of availability allows us to draw objective conclusions about the characteristics of those Web resources that are still available and correlate the features that distinguish them from disappeared or nonfunctional ones. These ten simple rules aid you in designing and maintaining a scientific Web resource that is available to anyone interested in using it.

## Rule 1: Plan Your Resource

As soon as you are seriously thinking about offering a Web resource to the general public, it is a good idea to lay down some ground rules. Clarify responsibilities in the processes of developing and maintaining the resource. Discuss these issues with the senior author or principal investigator, who is ultimately responsible for the availability of the resource. Read more about some ideas to manage responsibility in Rule 2.

Try to think of a good name that is not already taken and can be easily remembered. Changing the Web address of an existing resource is hard to do; it's better to start off with your own Internet domain name or a persistent URL. For the latter, the Online Computer Library Center offers a Persistent Uniform Resource Locator (PURL) for a changing Web address (for an overview, see [1]). It is essentially a transparent link to wherever your resource is currently hosted; its destination can be updated accordingly.

Some decisions early on can greatly impact the resource over its whole life cycle. Consider the level of service you want to offer. Is it a simple tool one step up from a command-line interface or a whole framework for large-scale analysis? How will users be able to access it? Read more about these options and how to make good use of the infrastructure available to you in Rule 4.

Throughout the life of your resource, there may be many different people involved in developing and maintaining it. Documentation is important for both developers and users of the resource. A scientific Web resource should be offered as open source software. Making your resource a software project at SourceForge.net, for instance, greatly facilitates development and maintenance. This also lets you keep an open channel of communications with your users, tell them about any major changes, and get their feedback to shape future developments.

Eventually, the resource may have outlived its usefulness. Read Rule 10 to find out when and how to shut down operations.

## Rule 2: Discuss Responsibilities

More than 58% of resources are developed entirely by researchers without a permanent position who will eventually move to another institution.

As a graduate student, involve your advisors early when you consider providing a Web resource. Chances are, they already know a way to share the work load. Discuss the issue of software maintenance, both for the time the original developers are still on site and for the time they have moved on. Do you want to take your work with you or leave it behind?

As an advisor, remember that this issue could come up, at the latest when your student leaves. As the senior author, solving such issues are your responsibility. Feel free to direct students towards using a certain software framework; creating such lab rules limits responsibility in a good way. You can even think of creating an intergenerational treaty for software maintenance among students in different years.

If your resource is used by collaborators and they think your program is valuable enough, you could convince them to take it over. The same is true for one of the following institutions: If your resource has a high impact and is useful to many people, you may be able to convince someone at the European Bioinformatics Institute (EBI), National Center for Biotechnology Information (NCBI), Netherlands Bioinformatics Centre (NBIC), or the PSU Galaxy instance to take over. Early decisions about the framework used can have a big impact later on.

## Rule 3: Know Your User Base

The most important component to consider is the Web resource audience. Come up with a use case: when and how will another researcher want to use what you are offering? When you know who you are developing for, many decisions become very straightforward. In our survey, we determined that 36% of Web resource providers think that only researchers with programming experience use their resource. If your audience can manage to run your application on their own computer, let them. It's harder to integrate a Web resource into a scripted workflow.

On the flip side, 64% of resources are also used by researchers without program-

**Citation:** Schultheiss SJ (2011) Ten Simple Rules for Providing a Scientific Web Resource. PLoS Comput Biol 7(5): e1001126. doi:10.1371/journal.pcbi.1001126

**Editor:** Philip E. Bourne, University of California San Diego, United States of America

**Published** May 26, 2011



**Funding:** This work was funded by the TL Stiftung Tuebingen, Germany: http://www.tl-stiftung.de/foerderpreise.php. The author holds a position at the Max Planck Society, Germany: http://fml.mpg.de. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

* E-mail: sebi@tuebingen.mpg.de

ming experience. They will appreciate a graphical user interface. If you know your users personally, they can give you ideas about how to make the interface fit their needs. Just watching collaborators or students use your software or programs like it will tell you a lot. Get users involved early and include them in the development process. As long as the Web resource is in use, you can solicit feedback from users and see if their needs have evolved (cf. Rule 7).

Constant monitoring of usage patterns and access statistics can be achieved by tracking who visits the Web resource page. If your institution is not already collecting these data from visitors, you can set up a free Web analytics tool within minutes.

Most scientists will come to your Web site via a search engine. Use the indexing power of the search engine spiders by putting, for example, the paper title, abstract, and keywords on the page. When you follow the tips about naming your resource in Rule 1, it should be easy to find.

## Rule 4: Use Services Available to You during Development

The finest way out of much of the strife with hosting and availability is to find someone else to take care of it. If you work on a larger campus or cooperate with someone at an institution that already runs several scientific Web resources, get in touch with the administrators to set up your tool on an established server. Such decisions can greatly influence the software development process. Be aware of the Web address you use to publish your resource. It's best to use a persistent URL or your own domain name for the resource to make sure it is always available under the published address (cf. Rule 1).

Estimate the number of potential simultaneous users. Together with the memory and compute time requirements, this will tell you about the kind of infrastructure you will have to provide to make the resource usable even with many queries coming in at the same time. In an age of high-throughput experiments, this can be a lot. To get an estimate on the number of simultaneous queries your setup can handle, you can perform a stress test, sending a high number of requests with a script from an external source.

If your requirements seem enormous, consider optimizing your program further and finding redundancies between individual queries that can be pre-computed and stored. You can also offer an interface to a cloud-computing on-demand resource, so users are paying for their own computing time. Providing your own large-scale computing infrastructure can be very costly.

You will have to think about a user interface for your resource. Here, an existing framework can save you a lot of time. Examples include Taverna [2], where you provide a description of the input and output in the Web resource description language. Your resource is accessed from a client workbench, in which users can connect your program's output to others to create workflows. It still runs on your own servers and you have to provide the necessary software infrastructure for that.

Galaxy [3] is a customizable workbench that you can download and run on your own Web server. It lets you integrate any command-line tool with a few lines of XML; moreover, it even lets you connect your own tools with the pre-packaged ones to create transparent workflows for your users. You don't need to think about file management and pretty user interfaces, and for those time-intensive jobs, you can easily connect your Galaxy instance to a compute cluster or even run it in the cloud.

If you want to build an interface from scratch, there are also frameworks that make this task easier. Aside from the classic Apache, SQL, and PHP combination, there are a few more modern alternatives: take a look at Ruby on Rails, Tomcat, Pyjamas, or CherryPy.

## Rule 5: Ensure Portability

Make sure that you can still install and run the software on another machine. If you want your software to be available three years from now, consider this strongly. Chances are that the server you are developing on will be replaced or software is updated, which often breaks the functionality. Ensuring portability also makes it easier for computational biologists to install your software locally. Ask a colleague to install the resource from scratch on another computer and you'll see where the pitfalls are.

A brute-force approach to portability is creating a virtual machine (VM). If you have a server where your resource runs just fine, back up its hard disk and restore it in a VM like VirtualBox. That way, you have a running version of your server in a single file. The VM approach is a steamroller tactic for resources with very intricate dependencies. This is a way to provide users with the necessary disk image to run your resource on the compute cloud. However, it is still advisable to provide information on how to set up your program from scratch. Together with source code comments and a high-level user manual, these three layers of documentation will ensure portability.

## Rule 6: Create an Open Source Project

Your source code should be public if the results are used in scientific publications. This is needed for reproducibility (read more about this in Rule 8).

To make your life easier, it is a good idea to place your source code in a repository such as SourceForge.net [4] or Bioinformatics.org [5]. Then you don't have to take care of version control and release issues and it's easier for collaborators to work together over distance. Most of these open source software project sites provide developers a means of communication both with each other and with end users. You can choose between mailing lists (with an online archive), a Web site forum, or an FAQ page.

Many scientists develop programs for one of the proprietary mathematical environments that require expensive licenses to run. If you are still in the planning stage, consider switching to an open source alternative. Your funding body may not be willing to pay for a score of licenses just for the users of your Web resource.

Using open source software, good source code documentation, and standard file formats will go a long way in making your software able to run on other computers (cf. Rules 5 and 7).

## Rule 7: Provide Ample Documentation and Listen to Feedback

A good first impression is very important for Web resources, too. It is crucial that first-time users feel welcome on your site. Provide good documentation and some short info about parameter settings, that is, accepted ranges and optional settings. Ideally, there is a one-click testing possibility with meaningful but easily understood example data. If the output of the example is well-defined, set it up to run periodically as a functional test, for instance during the build process.

Nothing teaches you about parameter settings, file formats, and the general purpose of a resource like a well-crafted demonstration of what it can do, for instance, in a video or screen cast. Many of these points are part of journals' instructions to authors and therefore required when submitting a research article about your Web resource.

A main complaint of the interviewed scientists about working resources was lack of documentation (41%). Beyond the reference to the paper to be cited when using the resource, you should include a brief overview of the resource's purpose, for what kinds of data it is applicable, and pointers to common pitfalls or preprocessing steps that are not so obvious. The latter is hard to imagine beforehand, so find out from users what they consider difficult.

It will be worth your while to set up a channel of communication with your users. Many source code repositories provide such functions (cf. Rule 6), which will save you a lot of time responding to frequent questions users ask about the resource. You can post announcements about maintenance, updates, and bug fixes, and best of all, experienced users often will be there to answer recurring questions raised by newbies, or you can refer them to the collective wisdom of the archives. It is also common practice to provide an e-mail address where the authors can be reached.

Make your life easier by providing a comprehensive error report option that users can click on when something fails, thereby e-mailing you all the information you need to find out what went wrong.

There are two more layers of documentation: in addition to the high-level help for end users, installation instructions will ensure portability, and good source code comments enable you to hand over maintenance responsibility to another developer, maybe even from the user community (cf. Rule 9).

## Rule 8: Facilitate Reproducibility

Reproducibility is always a topic of discussion in computational biology. When a user analyzes data with your Web resource, the results may end up in a research article. Therefore, all the steps needed to reproduce these results have to be documented entirely. In your output, provide users with details about the parameter settings they used, the version number, and information about the input data.

Everything to run the analysis again should be available to reviewers and readers. This includes the source code of the Web resource itself (cf. Rule 6).

It is good practice to make available older versions of the resource for purposes of reproducing results; at least boldly display the Web resource's current version number on the site and hints about how changes may affect the output.

If you change the server's behavior, your users have to know. Even if it is merely a bug fix, be sure to report it publicly in a place that will be noticed when using the server. Keep in mind that some users, for example, may have bookmarked the data submission page.

## Rule 9: Plan Ahead: Long-Term Maintenance

You will probably move to another place during your career. If you leave behind a Web resource, try to make the transition to the new maintainer as smoothly as possible. Ideally, a protocol has already been established during the planning phase (cf. Rule 1). In our survey, we found that more than 24% of Web resources will not be maintained after the original developers leave. Ultimately, it is the responsibility of the senior author of a publication to make sure that this does not happen, but it is a very important consideration for all authors of a Web resource publication.

Documentation of the source code and the installation process will greatly facilitate the transition to new maintainers. If there is no one in your old lab to take over, but the resource is still heavily used, you may be able to convince a current user or a collaborator to take over maintaining the resource. This will be even easier if the program is an open source software project, where a new developer can join at any time.

You may want to take your software with you and find a new home for it. In some circumstances, this requires you to change the Web resource's address. If your resource has been published in a journal, try contacting them and ask to have the link to your resource updated. Some journals may require a formal correction. Get your previous institution to link or forward to the new address from the old page for as long as possible. If you used a persistent URL, all you need to do is update the link (cf. Rule 1).

## Rule 10: Switch off an Unused Resource

During our study, we determined that, while a surprising number of Web resources are still available after a long time, they may not always work any longer. For users, this can be even more frustrating than an unavailable page.

If your resource is no longer under active development, chances are that it has outlived its usefulness after some years. After that, check to see if there is anyone still using it or if the original publication has been cited recently. This should be easy when you followed the advice about collecting statistics in Rule 3. If no one is using your resource any longer, release the source code one last time, and you're done.

If the resource is still useful to some researchers, try posting a notice on the site asking for someone to take over (cf. Rule 9). If all of that seems like too much work and the source code alone won't help anyone, consider creating a VM that runs the resource. When you still have access to the server, this can be done in a matter of hours.

By following these rules, your resource will have a long and productive life.


## Acknowledgments

The author would like to thank David J. Engel, Verena A. Kottler, Christoph U. Malisi, Gunnar Rätsch, and Eva-Maria Willing for critically reading the manuscript, and Jonas Behr, Regina Bohnert, Philipp Drewe, Andre Kahles, Vipin Thankam Sreedharan, and Christian Widmer for discussions. The comments from Philip E. Bourne and the anonymous reviewers were greatly appreciated and very helpful in improving this manuscript.


## References


1. Simonson J, Berleant D, Zhang X, Xie M, Vo H (1998) Version augmented URIs for reference permanencevia an Apache module design. Computer Networks and ISDN Systems 30: 337–345.
2. Hull D, Wolstencroft K, Stevens R, Goble C, Pocock MR, et al. (2006) Taverna: a tool for building and running workflows of services. Nucleic Acids Res 34: 729–732.
3. Goecks J, Nekrutenko A, Taylor J, The Galaxy Team (2010) Galaxy: a comprehensive approachfor supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11: R86. doi:10.1186/gb-2010-11-8-r86.
4. Geeknet, Inc (2010) SourceForge.net: open source software development. Available: http://sourceforge.net/. Accessed 8 April 2011.
5. Bioinformatics Organization (2010) Collaborative development environment. Available: http://www.bioinformatics.org/wiki/Hosting. Accessed 8 April 2011.

Editorial

# Ten Simple Rules for Reviewers

Philip E. Bourne*, Alon Korngreen

Last summer, the Student Council of the International Society for Computational Biology prompted an Editorial, "Ten Simple Rules for Getting Published" [1]. The interest in that piece (it has been downloaded 14,880 times thus far) prompted "Ten Simple Rules for Writing a Grant" [2]. With this third contribution, the "Ten Rules" series would seem to be established, and more rules for different audiences are in the making. *Ten Simple Rules for Reviewers* is based upon our years of experience as reviewers and as managers of the review process. Suggestions also came from PLoS staff and Editors and our research groups, the latter being new and fresh to the process of reviewing.

The rules for getting articles published included advice on becoming a reviewer early in your career. If you followed that advice, by working through your mentors who will ask you to review, you will then hopefully find these *Ten Rules for Reviewers* helpful. There is no magic formula for what constitutes a good or a bad paper—the majority of papers fall in between—so what do you look for as a reviewer? We would suggest, above all else, you are looking for what the journal you are reviewing for prides itself on. Scientific novelty—there is just too much "me-too" in scientific papers—is often the prerequisite, but not always. There is certainly a place for papers that, for example, support existing hypotheses, or provide a new or modified interpretation of an existing finding. After journal scope, it comes down to a well-presented argument and everything else described in "Ten Simple Rules for Getting Published" [1]. Once you know what to look for in a paper, the following simple reviewer guidelines we hope will be useful. Certainly (as with all *PLoS Computational Biology* material) we invite readers to use the PLoS eLetters feature to suggest their own rules and comments on this important subject.

## Rule 1: Do Not Accept a Review Assignment unless You Can Accomplish the Task in the Requested Timeframe—Learn to Say No

Late reviews are not fair to the authors, nor are they fair to journal staff. Think about this next time you have a paper under review and the reviewers are unresponsive. You do not like delays when it is your paper, neither do the authors of the paper you are reviewing. Moreover, a significant part of the cost of publishing is associated with chasing reviewers for overdue reviews. No one benefits from this process.

## Rule 2: Avoid Conflict of Interest

Reviews come in various forms—anonymous, open, and double-blind, where reviewers are not revealed to the authors and authors are not revealed to reviewers. Whatever the process, act accordingly and with the highest moral principles. The cloak of anonymity is not intended to cover scientific misconduct. Do not take on the review if there is the slightest possibility of conflict of interest. Conflicts arise when, for example, the paper is poor and will likely be rejected, yet there might be good ideas that you could apply in your own research, or, someone is working dangerously close to your own next paper. Most review requests first provide the abstract and then the paper only after you accept the review assignment. In clear cases of conflict, do not request the paper. With conflict, there is often a gray area; if you are in any doubt whatsoever, consult with the Editors who have asked you to review.

## Rule 3: Write Reviews You Would Be Satisfied with as an Author

Terse, ill-informed reviews reflect badly on you. Support your criticisms or praise with concrete reasons that are well laid out and logical. While you may not be known to the authors, the Editor knows who you are, and your reviews are maintained and possibly analyzed by the publisher's manuscript tracking system. Your profile as a reviewer is known by the journal—that profile of review quality as assessed by the Editor and of timeliness of review should be something you are proud of. Many journals, including this one, provide you with the reviews of your fellow reviewers after a paper is accepted or rejected. Read those reviews carefully and learn from them in writing your next review.

## Rule 4: As a Reviewer You Are Part of the Authoring Process

Your comments, when revisions are requested, should lead to a better paper. In extreme cases, a novel finding in a paper on the verge of rejection can be saved by (often) multiple rounds of revision based on detailed reviewers' comments and become highly cited. You are an unacknowledged partner in the success of the paper. It is always beneficial to remember that you are there to help the authors in their work, even if this means rejecting their manuscript.

**Citation:** Bourne PE, Korngreen A (2006) Ten simple rules for reviewers. PLoS Comput Biol 2(9): e110. DOI: 10.1371/journal.pcbi.0020110

**DOI:** 10.1371/journal.pcbi.0020110



Philip E. Bourne is a professor in the Department of Pharmacology, University of California San Diego, La Jolla, California, United States of America, and is Editor-in-Chief of *PLoS Computational Biology.* Alon Korngreen is a Lecturer in the Mina and Everard Faculty of Life Sciences and the Leslie and Susan Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel.

* To whom correspondence should be addressed. E-mail: bourne@sdsc.edu

## Rule 5: Be Sure to Enjoy and to Learn from the Reviewing Process

Peer review is an important community service and you should participate. Unfortunately, the more you review, in all likelihood the more you will be asked to review. Often you will be asked to review boring papers that are of no interest to you. While it is important to serve as a reviewer, only accept papers in which you are keenly interested, because either they are close to your area of research or you feel you can learn something. You might say, should I not know the work very well to be a reviewer? Often a perspective from someone in a slightly different area can be very effective in improving a paper. Do not hesitate to indicate to the Editor the perspective that you can bring to a paper (see Rule 10); s/he can then decide how to weigh your review. Editors would of course like to see you review papers even if you are not particularly interested in them, but the reality is that good reviewers must use their reviewing time wisely.

## Rule 6: Develop a Method of Reviewing That Works for You

This may be different for different people. A sound approach may be to read the manuscript carefully from beginning to end before considering the review. This way you get a complete sense of the scope and novelty of the work. Then read the journal's Guide to Authors, particularly if you have not published in the journal yourself, or if the paper is a particular class of article with which you are not overly familiar, a review for example. With this broad background, you can move to analyzing the paper in detail, providing a summary statement of your findings as well as detailed comments. Use clear reasoning to justify each criticism, and highlight the good points about the work as well as the weaker points. Including citations missed by the author (not your own) is often a short but effective way to help improve a paper. A good review touches on both major issues and minor details in the manuscript.

## Rule 7: Spend Your Precious Time on Papers Worthy of a Good Review

The publish-or-perish syndrome leads to many poor papers that may not be filtered out by the Editors prior to sending it out for review. Do not spend a lot of time on poor papers (this may not be obvious when you take on the paper by reading only the abstract), but be very clear as to why you have spent limited time on the review. If there are positive aspects of a poor paper, try to find some way of encouraging the author while still being clear on the reasons for rejection.

## Rule 8: Maintain the Anonymity of the Review Process if the Journal Requires It

Many of us have received reviews where it is fairly obvious who reviewed the work, sometimes because they suggest you cite their work. It is hard to maintain anonymity in small scientific communities, and you should reread your review to be sure it does not endanger the anonymity if anonymous reviews are the policy of the journal. If anonymity is the rule of the journal, do not share the manuscript with colleagues unless the Editor has given the green light. Anonymity as a journal policy is rather a religious rule—people are strongly for and against. Conform strictly to the policy defined by the journal asking you to review.

## Rule 9: Write Clearly, Succinctly, and in a Neutral Tone, but Be Decisive

A poorly written review is as bad as a poorly written paper (see Rule 3). Try to be sure the Editors and the authors can understand the points you are making. A point-by-point critique is valuable since it is easy to read and to respond to. For each point, indicate how critical it is to your accepting the paper. If English is not your strong point, have someone else read the paper and the review, but without violating other rules, particularly Rule 2. Further, as passionate as you might be about the subject of the paper, do not push your own opinion or hypotheses. Finally, give the Editors a clear answer as to your recommendation for publication. Reviewers frequently do not give a rating even when requested. Provide a rating—fence-sitting prolongs the process unnecessarily.

## Rule 10: Make Use of the "Comments to Editors"

Most journals provide the opportunity to send comments to the Editors, which are not seen by the authors. Use this opportunity to provide your opinion or personal perspective of the paper in a few clear sentences. However, be sure those comments are clearly supported by your review—do not leave the Editor guessing with comments like "this really should not be published" if your review does not strongly support that statement. It is also a place where anonymity can be relaxed and reasons for decisions made clearer. For example, your decision may be based on other papers you have reviewed for the journal, which can be indicated in the Editor-only section. It is also a good place to indicate your own shortcomings, biases, etc., with regard to the content of the paper (see Rule 5). This option is used too infrequently and yet can make a great deal of difference to an Editor trying to deal with a split decision. ■

**References**

1. Bourne PE (2005) Ten simple rules for getting published. PLoS Comput Biol 1 (5): DOI: 10.1371/journal.pcbi.0010057
2. Bourne PE, Chalupa LM (2006) Ten simple rules for getting grants. PLoS Comput Biol 2 (2): DOI: 10.1371/journal.pcbi.0020012