

Meta Review

This manuscript was automatically generated from greenelab/meta-review@bb524ab on September 30, 2017.

Authors

- **Daniel S. Himmelstein**

[0000-0002-3012-7446](#) · [dhimmel](#) · [dhimmel](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552

- **Casey S. Greene**

[0000-0001-8713-9213](#) · [cgreene](#) · [GreeneScientist](#)

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania · Funded by GBMF4552

- **Anthony Gitter**

[0000-0002-5324-9833](#) · [agitter](#) · [anthonygitter](#)

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison and Morgridge Institute for Research · Funded by NIH U54AI117924

Abstract

Main text

Introduction

Open research – which includes sharing code, data, and manuscripts – benefits the researchers who engage in open practices [1], their scientific peers, and the public.

`TODO: more references needed` Here we describe the benefits of writing review articles openly, where the planning, organizing, writing, and editing occur collaboratively in a public forum where participants are free to join as they wish. Reviews presenting the state of the art in a scientific field are often prepared by a single research group or a small team of colleagues. In contrast, broadly opening the writing process to anyone engaged in the topic can help maximize the review's value by facilitating the representation of diverse opinions and the broad coverage of relevant research. Review authors can engage with the authors of original research to clarify their methods and results and present them accurately (for example, [2]). `TODO: need archival issue link` In addition, discussing manuscripts in the open provides one form of pre- and post-publication peer review `TODO: define this or provide a reference?`, incentivizing the reviews with potential manuscript authorship. However, inviting wide authorship brings many technical and social challenges such as how to fairly distribute credit, coordinate the scientific content, and collaboratively manage extensive reference lists.

We present solutions to these challenges based on our recent experience leading a collaborative review “Opportunities And Obstacles For Deep Learning In Biology And Medicine” [3]. Our review attracted 27 authors from 20 different institutions who were not determined in advance.

`TODO: confirm institution count` We wrote entirely in the open without restrictions on who was welcome to contribute. Although we requested that some authors participate for their specific expertise, most discovered the manuscript organically through conferences or social media and independently decided to contribute. `TODO: confirm "most"` To coordinate this effort, we developed a manuscript writing process using the Markdown language, the GitHub software development platform [4], and our new Manubot tool [5,6] for automating manuscript generation.

Collaborative writing and reviewing

There are many existing collaborative writing platforms ranging from rich text editors, which support Microsoft Word documents or similar formats, to LaTeX-based systems for more technical writing [7–9]. These platforms ideally offer version control, multiple permission levels, or other functionality to support multi-author document editing. Although they work well for editing, they lack sufficient features for managing a collaborative manuscript and attributing precise credit, which are important for open writing.

We adopted standard software development strategies in order to enable any contributor to edit any part of the manuscript but enforce discussion and review of all proposed changes. The GitHub platform provided support for organizing and editing the manuscript. We used GitHub *issues* for organization, opening a new issue for each paper under consideration. Within the issue, contributors summarized the research, discussed it (sometimes with the original authors), and assessed its relevance to the review. Issues also served as an open to-do list and a forum for debating the main message, themes, and topics of the review.

GitHub and the underlying git version control system [10,11] also structured the writing process. The official version of the manuscript is *forked* by individual contributors. A contributor then adds and revises files, grouping these changes into *commits*. When the changes are ready to be reviewed, the series of commits are submitted as a *pull request* through GitHub, which notifies other authors of the pending changes. GitHub's review interface allows anyone to comment on the changes, globally or at specific lines, asking questions or requesting modifications as depicted in [12]. Conversations during review can reference other pull requests, issues, or authors, linking the relevant people and content. Reviewing batches of revisions that focus on a single theme is more efficient than independently discussing isolated comments and edits, and it helps maintain consistent content and tone across different authors and reviewers. Once all requested modifications are made, the manuscript maintainers, a subset of authors with elevated GitHub permissions, formally approve the pull request and merge the changes into the official version.

`TODO: need a figure with a flowchart showing this process` The entire process can be orchestrated through GitHub with a web browser if a contributor does not want to use a git client on their own computer.

We found that this workflow was an effective compromise between fully unrestricted editing and a more heavily-structured approach that limits the authors or the sections they can edit. In addition, authors are associated with their commits, which makes it easy for contributors to receive credit for their work and helps prevent ghostwriting [13]. The GitHub contributors page summarizes all edits and commits from each author, providing aggregated information that is not available on other collaborative writing platforms. Because our writing process, like others backed by the open git version control system [8,9], tracks the complete commit history, it also enables detailed

retrospective contribution analysis.

TODO: confirm Overleaf and Authorea provide this type of git integration versus something more coarse

Authorship

To determine authorship we followed the International Committee of Medical Journal Editors (ICMJE) guidelines and used GitHub to track contributions. ICMJE recommends authors substantially contribute to, draft, approve, and agree to be accountable for the manuscript [14]. We acknowledged other contributors who did not meet all four criteria, including contributors who provided text but did not review and approve the complete manuscript.

TODO: In Discussion, mention the challenge of determining contributions of potential authors who worked with someone else on the

Although these criteria provided a straightforward, equitable way to determine who would be an author, they did not produce a traditionally ordered author list. In biomedical journals, the convention is that the first and last authors made the most substantial contributions to the manuscript. This convention can be difficult to reconcile in a collaborative effort. Using git, we could quantify the number of commits each author made or the number of sentences an author wrote or edited, but these metrics discount intellectual contributions such as discussing primary literature and reviewing pull requests. However, there is no objective system to compare and weight the different types of contributions and produce an ordered author list.

To address this issue, we generalized the concept of “co-first” authorship, in which two or more authors are denoted as making equal contributions to a paper. We defined four types of contributions [3], from major to minor, and reviewed the GitHub discussions and commits to assign authors to these categories. A randomized algorithm then arbitrarily ordered authors within each contribution category, and we combined the category-specific author lists to produce a traditional ordering. The randomization procedure was shared with the authors in advance (pre-registered) and run in a deterministic manner. Given the same author contributions, it always produced the same ordered author list. We annotated the author list to indicate that author order was partly randomized and emphasize that the order did not indicate one author contributed more than another from the same category.

TODO: In Discussion, present alternative author ordering strategies and literature on contribution in collaborative projects

Discussion

Many others have embraced open science principles and piloted open approaches toward drug discovery [15,16], data management [17–19], and manuscript review [20].

TODO: need help deciding what related topics to include here and which references to use, these are arbitrary examples

TODO: more ideas in doi:10.7287/peerj.preprints.2711v2 Several of these open science efforts are GitHub-based like our collaborative writing process. The Journal of Open Source Software [21] and ReScience [22] journals rely on GitHub for peer review and hosting.

TODO: describe Manubot related work here? [23,24]

There are potential limitations of our GitHub-based approach. Because our review manuscript pertained to a computational topic, most of the authors had computational backgrounds, including previous experience with version control workflows and GitHub. In other disciplines, collaborative writing via GitHub and Manubot could present a steeper barrier to entry and deter participants. In addition, git carefully tracks all revisions to the manuscript text but not the surrounding conversations that take place through GitHub issues and pull requests. These discussions must be archived to ensure that important decisions about the manuscript are preserved and authors receive credit for intellectual contributions that are not directly reflected in the manuscript’s text. GitHub supports programmatic access to issues, pull requests, and reviews so tracking these conversations is feasible in the future.

TODO: cgreene paragraph on diversity, open science culture, possibly author disagreements and conduct

Open writing presents new opportunities for scholarly communication.

TODO: reference "paper of the future"? arXiv:1601.02927 doi:10.22541/au.149693987.70506124 doi:10.22541/au.148769949.92783646 ht

Though it is still valuable to have versioned drafts of a review manuscript with digital identifiers, journal publication may not be the terminal endpoint for collaborative manuscripts. After releasing the first version of our collaborative review [3], six new authors have contributed text and existing authors continue to discuss new literature, creating a living document [4].

TODO: update new author count before submitting The Manubot system can also facilitate open research [25] in addition to review articles.

TODO: get permission and add https://slochowier.github.io/nonequilibrium-barrier/ https://zietzm.github.io/Vagelos2017/

Our process represents an early step toward open massively collaborative reviews, and there are certainly aspects that can be improved. We invite the scientific community to adapt and build upon our experience and open software.

Acknowledgements

TODO: deep review authors for support in testing this process

TODO: manubot-rootstock contributors

References

References

1. How open science helps researchers succeed

Erin C McKiernan, Philip E Bourne, C Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, Brian A Nosek, Karthik Ram, Courtney K Soderberg, ... Tal Yarkoni
eLife (2016-07-07) <https://doi.org/10.7554/elife.16800>

2. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology · Issue #213 · greenelab/deep-review

greenelab
GitHub <https://github.com/greenelab/deep-review/issues/213>

3. Opportunities And Obstacles For Deep Learning In Biology And Medicine

Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Wei Xie, Gail L. Rosen, ... Casey S. Greene
Cold Spring Harbor Laboratory (2017-05-28) <https://doi.org/10.1101/142760>

4. greenelab/deep-review

greenelab
GitHub <https://github.com/greenelab/deep-review>

5. greenelab/manubot-rootstock GitHub repository

Daniel Himmelstein
GitHub (2017) <https://github.com/greenelab/manubot-rootstock>

6. greenelab/manubot

dhimmel
GitHub <https://github.com/greenelab/manubot>

7. Scientific writing: the online cooperative

Jeffrey M. Perkel
Nature (2014-10-01) <https://doi.org/10.1038/514127a>

8. Overleaf: Real-time Collaborative Writing and Publishing Tools with Integrated PDF

Preview@overleaf <https://www.overleaf.com/>

9. Authorea@authorea <https://www.authorea.com/>

10. A Quick Introduction to Version Control with Git and GitHub

John D. Blischak, Emily R. Davenport, Greg Wilson
PLOS Computational Biology (2016-01-19) <https://doi.org/10.1371/journal.pcbi.1004668>

11. Ten Simple Rules for Taking Advantage of Git and GitHub

Yasset Perez-Riverol, Laurent Gatto, Rui Wang, Timo Sachsenberg, Julian Uszkoreit, Felipe da Veiga Leprevost, Christian Fufezan, Tobias Bernert, Stephen J. Eglen, Daniel S. Katz, ... Juan Antonio Vizcaíno
PLOS Computational Biology (2016-07-14) <https://doi.org/10.1371/journal.pcbi.1004947>

12. Opportunities And Obstacles For Deep Learning In Biology And Medicine

Johnny Israeli
Medium (2017-05-31) <https://medium.com/towards-data-science/opportunities-and-obstacles-for-deep-learning-in-biology-and-medicine-6ec914fe18c2>

13. What Should Be Done To Tackle Ghostwriting in the Medical Literature?

Peter C Gøtzsche, Jerome P Kassirer, Karen L Woolley, Elizabeth Wager, Adam Jacobs, Art Gertel, Cindy Hamilton
PLoS Medicine (2009-02-03) <https://doi.org/10.1371/journal.pmed.1000023>

14. ICMJE | Recommendations | Defining the Role of Authors and

Contributors<http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>

15. Systematic integration of biomedical knowledge prioritizes drugs for repurposing

Daniel S. Himmelstein, Antoine Lizée, Christine Hessler, Leo Brueggeman, Sabrina L. Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, Sergio E. Baranzini
Cold Spring Harbor Laboratory (2016-11-14) <https://doi.org/10.1101/087619>

16. An open source pharma roadmap

Manica Balasegaram, Peter Kolb, John McKew, Jaykumar Menon, Piero Olliaro, Tomasz Sablinski, Zakir Thomas, Matthew H. Todd, Els Torrelee, John Wilbanks
PLOS Medicine (2017-04-18) <https://doi.org/10.1371/journal.pmed.1002276>

17. The Open Knowledge Foundation: Open Data Means Better Science

Jennifer C. Molloy

PLoS Biology (2011-12-06) <https://doi.org/10.1371/journal.pbio.1001195>

18. The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, ... Barend Mons

Scientific Data (2016-03-15) <https://doi.org/10.1038/sdata.2016.18>

19. First, design for data sharing

John Wilbanks, Stephen H Friend

Nature Biotechnology (2016-03-03) <https://doi.org/10.1038/nbt.3516>

20. A multi-disciplinary perspective on emergent and future innovations in peer review

Jonathan P. Tennant, Jonathan M. Dugan, Daniel Graziotin, Damien C. Jacques, François Waldner, Daniel Mietchen, Yehia Elkhatib, Lauren B. Collister, Christina K. Pikas, Tom Crick, ... Julien Colomb

F1000Research (2017-07-20) <https://doi.org/10.12688/f1000research.12037.1>

21. Journal of Open Source Software (JOSS): design and first-year review

Arfon M Smith, Kyle E Niemeyer, Daniel S Katz, Lorena A Barba, George Githinji, Melissa Gymrek, Kathryn D Huff, Christopher R Madan, Abigail Cabunoc Mayes, Kevin M Moerman, ... Jacob T Vanderplas

arXiv (2017-07-07) <https://arxiv.org/abs/1707.02264v1>

22. Sustainable computational science: the ReScience initiative

Nicolas P. Rougier, Konrad Hinsén, Frédéric Alexandre, Thomas Arildsen, Lorena Barba, Fabien C. Y. Benureau, C. Titus Brown, Pierre de Buyl, Ozan Caglayan, Andrew P. Davison, ... Tiziano Zito

arXiv (2017-07-14) <https://arxiv.org/abs/1707.04393v1>

23. Formatting Open Science: agilely creating multiple document formats for academic manuscripts with Pandoc Scholar

Albert Krewinkel, Robert Winkler

PeerJ Computer Science (2017-05-08) <https://doi.org/10.7717/peerj-cs.112>

24. ewanmellor/gh-publisher

ewanmellor

GitHub <https://github.com/ewanmellor/gh-publisher>

25. Sci-Hub provides access to nearly all scholarly literature

Daniel S Himmelstein, Ariel R Romero, Stephen R McLaughlin, Bastian Greshake Tzovaras, Casey S Greene

PeerJ Preprints (2017-07-20) <https://doi.org/10.7287/peerj.preprints.3100v1>