Meta Review

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

Offspring of the Deep Review. We cite the Deep Review like [1] or [1]. The source repository is [2].

Main text

Introduction

Open research – which includes sharing code, data, and manuscripts – benefits the researchers who engage in open practices [3], 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, [4]). `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 potential solutions to these challenges based on our recent experience leading a collaborative review "Opportunities And Obstacles For Deep Learning In Biology And Medicine" [1]. 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 [5], and our new Manubot tool [2,6] 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] 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 [11]. 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 [12]. 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 [13]. 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 [1], 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
```

Acknowledgements

References

1. Ching T *et al.* 2017 Opportunities And Obstacles For Deep Learning In Biology And Medicine. See https://doi.org/10.1101/142760.

2. Himmelstein D. 2017 greenelab/manubot-rootstock GitHub repository. *GitHub*. See https://github.com/greenelab/manubot-rootstock.

3. McKiernan EC *et al.* 2016 How open science helps researchers succeed. *eLife* **5**. See https://doi.org/10.7554/elife.16800.

4. greenelab. In press. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology · Issue #213 · greenelab/deep-review. *GitHub*. See https://github.com/greenelab/deep-review/issues/213.

5. greenelab. In press. greenelab/deep-review. *GitHub*. See https://github.com/greenelab/deep-review.

6. dhimmel. In press. greenelab/manubot. *GitHub*. See https://github.com/greenelab/manubot.

7. Perkel JM. 2014 Scientific writing: the online cooperative. *Nature* **514**, 127–128. See https://doi.org/10.1038/514127a.

8. In press. Overleaf: Real-time Collaborative Writing and Publishing Tools with Integrated PDF Preview. *@overleaf*. See https://www.overleaf.com/.

9. In press. Authorea. *@authorea*. See https://www.authorea.com/.

10. Blischak JD, Davenport ER, Wilson G. 2016 A Quick Introduction to Version Control with Git and GitHub. *PLoS Comput Biol* **12**, e1004668. See https://doi.org/10.1371/journal.pcbi.1004668.

11. Israeli J. 2017 Opportunities And Obstacles For Deep Learning In Biology And Medicine. *Medium*. See https://medium.com/towards-data-science/opportunities-and-obstacles-for-deep-learning-in-biology-and-medicine-6ec914fe18c2.

12. Gøtzsche PC, Kassirer JP, Woolley KL, Wager E, Jacobs A, Gertel A, Hamilton C. 2009 What

Should Be Done To Tackle Ghostwriting in the Medical Literature? *PLoS Med* **6**, e1000023. See https://doi.org/10.1371/journal.pmed.1000023.

13. In press. ICMJE | Recommendations | Defining the Role of Authors and Contributors. See http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html.