

Open collaborative writing with Manubot

This manuscript ([permalink](#)) was automatically generated from [greenelab/meta-review@12fd168](#) on October 21, 2018.

Authors

- **Daniel S. Himmelstein**

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

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania

- **David R. Slochower**

 [0000-0003-3928-5050](#) ·  [slochower](#) ·  [drslochower](#)

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego

- **Venkat S. Malladi**

 [0000-0002-0144-0564](#) ·  [vsmalladi](#) ·  [katatonikkat](#)

Department of Bioinformatics, University of Texas Southwestern Medical Center; Bioinformatics Core Facility,
University of Texas Southwestern Medical Center

- **Casey S. Greene**

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

Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania

- **Anthony Gitter**

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

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison; Morgridge Institute for
Research

Abstract

Open, collaborative research is a powerful paradigm that can immensely strengthen the scientific process by integrating broad and diverse expertise. However, traditional research and multi-author writing processes break down at scale. New tools and workflows that rely on automation can ensure correctness and fairness in massively collaborative research. We present techniques for overcoming challenges of open research, with special emphasis on manuscript writing. These include approaches for managing distributed authors and our new software, named Manubot, for automating citation and many other aspects of manuscript building.

Introduction

The internet enables science to be shared in real-time at a low cost to a global audience. This development has decreased the barriers to making science open, while supporting new massively collaborative models of research. However, the scientific community requires tools whose workflows encourage openness. Manuscripts are the cornerstone of scholarly communication, but drafting and publishing manuscripts has traditionally relied on proprietary or offline tools that do not support *open scholarly writing*, in which anyone is able to contribute and the contribution history is preserved and public. We introduce [Manubot](#), a new tool and infrastructure for authoring scholarly manuscripts in the open, and report how it was instrumental for the collaborative project that led to its creation.

Based on our experience leading a recent open review [1], we discuss the advantages and challenges of open collaborative writing, a form of crowdsourcing [2]. Our review manuscript [3] was code-named the Deep Review and surveyed deep learning's role in biology and precision medicine, a research area undergoing explosive growth. We initiated the Deep Review by creating a GitHub repository (<https://github.com/greenelab/deep-review>) to coordinate and manage contributions. GitHub is a platform designed for collaborative software development that is adaptable for collaborative writing. From the start, we made the GitHub repository public under a [Creative Commons Attribution License](#). We encouraged anyone interested to contribute by proposing changes or additions. Although we invited some specific experts to participate, most authors discovered the manuscript organically through conferences or social media, deciding to contribute without solicitation. In total, the Deep Review attracted 36 authors, who were not determined in advance, from 20 different institutions.

The Deep Review and other studies that subsequently adopted the Manubot platform were unequivocal successes bolstered by the collaborative approach. However, inviting wide authorship brought many technical and social challenges such as how to fairly distribute credit, coordinate the scientific content, and collaboratively manage extensive reference lists. The manuscript writing process we developed using the Markdown language, the GitHub platform, and our new Manubot tool for automating manuscript generation addresses these challenges.

Contribution workflow

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 technical writing [4] such as [Overleaf](#) and [Authorea](#). These platforms ideally offer version control, multiple permission levels, or other functionality to support multi-author document editing. Although they work well for editing text, they lack sufficient features for managing a collaborative manuscript and attributing precise credit, which are important for open writing (Table 1).

Table 1: Collaborative writing platforms. A summary of features that differentiate Manubot from existing collaborative writing platforms. We assessed features on June 15, 2018 using the free version of each platform. Some platforms offer additional features through a paid subscription or software. ¹Additional functionality, such as bibliography management, is available by editing the Word document stored in OneDrive with the paid Word desktop application. ²Conversations about modifications take place on the document as comments, annotations, or unsaved chats. There is no integrated forum for discussing and editing revisions. ³In some circumstances, Overleaf git commits are not modular. Edits made by distinct authors may be attributed to a single author.

Feature	Manubot	Authorea + BibTeX	Overleaf v1 + BibTeX	Google Docs + Paperpile	Word Online ¹	Markdown on GitHub
Multi-author editing	Yes	Yes	Yes	Yes	Yes	Yes
Propose changes	Yes	No	No	Yes	No	Yes
Continuous integration testing	Yes	No	No	No	No	No
Multi-participant conversation for changes	Yes	No ²	No ²	No ²	No ²	Yes
Character-level provenance for text	Yes	No (versions tracked by day)	No ³	Requires manual inspection of history	Not after changes are accepted	Yes
Bibliography management	Yes	Yes	Yes	Yes	No, requires the Word desktop application	No
Cite by common identifiers	Yes	No	No	No	No	No

Feature	Manubot	Authorea + BibTeX	Overleaf v1 + BibTeX	Google Docs + Paperpile	Word Online ¹	Markdown on GitHub
Editing software	Any text editor	Web interface	Web interface	Web interface	Web interface	Any text editor
Document format	Markdown	LaTeX	LaTeX	Proprietary	Proprietary	Markdown
Templating	Yes	Yes	Yes	No	No	No
Technical expertise required	Yes	Yes	Yes	No	No	Yes

In our workflow, we adopt standard software development strategies that enable any contributor to edit any part of the manuscript but enforce discussion and review of all proposed changes. The GitHub platform supports organizing and editing the manuscript. We use GitHub *issues* for organization, opening a new issue for each discussion topic. For example, in a review manuscript like the Deep Review, this includes each primary paper under consideration. Within a paper's issue, contributors summarize the research, discuss it (sometimes with participation from the original authors), and assess its relevance to the review. Issues also serve 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 [5,6] also structure 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 7. Conversations during review can reference other pull requests, issues, or authors, linking the relevant people and content, as illustrated in Figure 1. Reviewing batches of revisions that focus on a single theme is more efficient than independently discussing isolated comments and edits and 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. The process of writing and revising material can be orchestrated through GitHub with a web browser or a local text editor.

issues. The contributor made the pull request, and two maintainers and another participant made recommendations. After four rounds of reviews and pull request edits, a maintainer merged the changes.

We found that this workflow was an effective compromise between fully unrestricted editing and a more heavily-structured approach that limited the authors or the sections they could 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 [8]. Figure 2 and the GitHub [contributors page](#) summarize all edits and commits from each author, providing aggregated information that is not available on other collaborative writing platforms. Because the Manubot writing process tracks the complete history through git commits, it enables detailed retrospective contribution analysis.

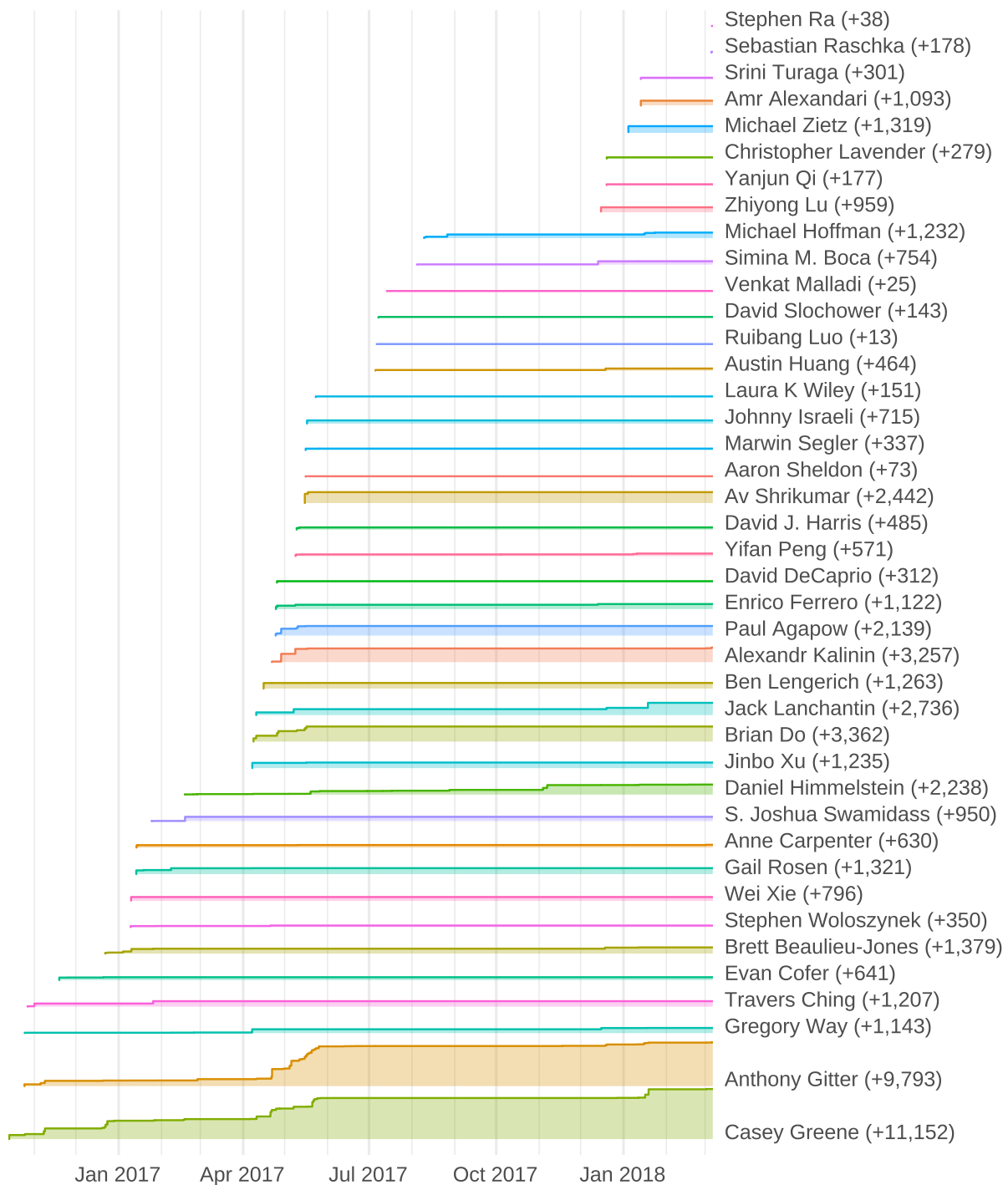


Figure 2: Deep Review contributions by author over time. The total words added to the Deep Review by each author is plotted over time (final values in parentheses). These statistics were extracted from git commit diffs of the manuscript's Markdown source. This figure reveals the composition of written contributions to the manuscript at every point in its history. The Deep Review was initiated in August 2016, and the first complete manuscript was released as a preprint [9] in May 2017. While the article was under review, we continued to maintain the project and accepted new contributions. The preprint was updated in January 2018, and the article was accepted by the

journal in March 2018 [3]. As of June 15, 2018, the Deep Review repository accumulated 755 git commits, 315 merged pull requests, 537 issues, and 616 GitHub stars.

Manubot

Manubot is a system for writing scholarly manuscripts via GitHub that is built upon our [Python package](#) of the same name. With Manubot, manuscripts are written as plain-text Markdown files, which is well suited for version control using git. The Markdown standard itself provides limited yet crucial formatting syntax, including the ability to embed images and format text via bold, italics, hyperlinks, headers, inline code, codeblocks, blockquotes, and numbered or bulleted lists. In addition, Manubot relies on extensions from [Pandoc Markdown](#) to enable citations, tables, captions, and equations specified using the popular TeX math syntax.

Manubot includes an additional layer of citation processing, currently unique to the system. All citations point to a standard identifier, for which Manubot automatically retrieves bibliographic metadata. Table 2 presents the supported identifiers and example citations before and after Manubot processing. Authors can optionally define citation tags to provide short readable alternatives to the citation identifiers. Metadata is exported to [Citation Style Language \(CSL\) JSON Data Items](#), an open standard that is widely supported by reference managers [10,11]. However, sometimes external resources provide Manubot with invalid CSL Data, which can cause errors with downstream citation processors, such as [pandoc-citeproc](#). Therefore, Manubot removes invalid fields according to the [CSL Data specification](#). In cases where automatic retrieval of metadata fails or produces incorrect references — which is most common for URL citations — users can manually provide the correct CSL JSON.

Table 2: Citation types supported by Manubot. Authors may optionally map a named tag to one of the other supported identifier types. In this example, the tag `avasthi-preprints` represents the DOI identifier `doi:10.7554/eLife.38532`.

Identifier	Metadata source	Example citation	Processed citation
Digital Object Identifier (DOI)	DOI Content Negotiation	<code>doi:10.1098/rsif.2017.0387</code>	[3]
PubMed Identifier (PMID)	NCBI's E-utilities	<code>pmid:25851694</code>	[12]
PubMed Central Identifier (PMCID)	NCBI's Citation Exporter	<code>pmcid:PMC4719068</code>	[2]
arXiv identifier	arXiv API	<code>arxiv:1502.04015v1</code>	[13]
URL	Greycite [14]	<code>url:https://lgatto.github.io/open-and-open/</code>	[15]
Tag	Source for tagged identifier	<code>tag:avasthi-preprints</code>	[16]

Manubot formats bibliographies according to a [CSL](#) style specification. Styles define how references are constructed from bibliographic metadata, controlling layout details such as the maximum number of authors to list per reference. Manubot's default style emphasizes titles and electronic (rather than print) identifiers and applies numeric-style citations [17]. Alternatively, users can also choose from thousands of [predefined styles](#) or build their own [18]. As a result, adopting the specific bibliographic format required by a journal usually just requires specifying the style's source URL in the Manubot configuration.

Manubot uses [Pandoc](#) to convert manuscripts from Markdown to HTML, PDF, and optionally DOCX outputs. Pandoc supports conversion between additional formats — such as LaTeX, AsciiDoc, EPUB, and JATS — offering Manubot users broad interoperability. [Journal Article Tag Suite](#) (JATS) is a standard XML format for scholarly articles that is used by publishers, archives, and text miners [19,20,21]. Pandoc's JATS support provides an avenue to integrate Manubot with

the larger JATS ecosystem. For now, the primary Manubot output is HTML intended to be viewed in a web browser.

Manubot performs continuous publication: every update to a manuscript's source is automatically reflected in the online outputs. The approach uses continuous integration (CI) [22,23,24], specifically via [Travis CI](#), to monitor changes. When changes occur, the CI service attempts to generate an updated manuscript. If this process is error free, the CI service timestamps the manuscript and uploads the output files to the GitHub repository. Because the HTML manuscript is hosted using [GitHub Pages](#), the CI service automatically deploys the new manuscript version when it pushes the updated outputs to GitHub. Using CI to build the manuscript automatically catches many common errors, such as misspelled citations, invalid formatting, or misconfigured software dependencies.

To illustrate, the source GitHub repository for this article is <https://github.com/greenelab/meta-review>. When this repository changes, Travis CI [rebuilds](#) the manuscript. If successful, the output is deployed back to GitHub (to dedicated `output` and `gh-pages` branches). As a result, <https://greenelab.github.io/meta-review> stays up to date with the latest HTML manuscript. Furthermore, versioned URLs, such as <https://greenelab.github.io/meta-review/v/4b6396bcefd1b9c7ddf39c1d3f0b3eab2dd63f31/>, provide access to previous manuscript versions.

The idea of the “priority of discovery” is important to science, and Vale and Hyman discuss the importance of both disclosure and validation [25]. In their framework, disclosure occurs when a scientific output is released to the world. However, for a manuscript that is shared as it is written, being able to establish priority could be challenging. Manubot supports [OpenTimestamps](#) to timestamp the HTML and PDF outputs on the Bitcoin blockchain. This procedure allows one to retrospectively prove that a manuscript version existed prior to its blockchain-verifiable timestamp [13,26,27,28]. Timestamps protect against attempts to rewrite a manuscript's history and ensure accurate histories, potentially alleviating certain authorship or priority disputes. Because all bitcoin transactions compete for limited space on the blockchain, the fees required to send a single transaction can be high. OpenTimestamps avoids this fee by encoding many timestamps into a single Bitcoin transaction [29]. There can be a lag of a few hours before the transaction is made, which is suitable for the purposes of scientific writing.

Manubot and its dependencies are free of charge and largely open source. It does rely on gratis services from two proprietary platforms: GitHub and Travis CI. Fortunately, lock-in to these services is minimal, and several substitutes already exist. Manubot provides a substantial step towards end-to-end document reproducibility, where every figure or piece of data in a manuscript can be traced back to its origin [30] and is well suited for preserving provenance. For example, figures can be specified using versioned URLs that refer to the code that created them. In addition, manuscripts can be templated, so that numerical values or tables are inserted directly from the repository that

created them. An [example repository](#) demonstrates Manubot's features and serves as a template for users to write their own manuscript with Manubot.

Since its creation to facilitate the Deep Review, Manubot has been used to write a variety of scholarly documents. The [Sci-Hub Coverage Study](#) — performed openly on GitHub from its inception — investigated Sci-Hub's repository of pirated articles [31]. Sci-Hub [reviewed](#) the initial preprint from this study in a series of tweets, pointing out a major error in one of the analyses. Within hours, the authors used Markdown's strikethrough formatting in Manubot to cross-out the errant sentences ([commit](#), [versioned manuscript](#)), thereby alerting readers to the mistake and preventing further propagation of misinformation. One month later, a larger [set of revisions](#) explained the error in more detail and was included in a second version of the preprint. As such, continuous publishing via Manubot helped the authors address the error without delay, while retaining a public version history of the process. This Sci-Hub Coverage Study preprint was the [most viewed](#) 2017 *PeerJ Preprint*, while the Deep Review was the most viewed 2017 *bioRxiv* preprint [32]. Hence, in Manubot's first year, two of the most popular preprints were written using its collaborative, open, and review-driven authoring process.

Additional research studies in progress are being authored using Manubot, spanning the fields of [genomics](#), [climate science](#), and [data visualization](#). Manubot is also being used for documents beyond traditional journal publications, such as [grant proposals](#), [progress reports](#), [undergraduate research reports](#) [33], [literature reviews](#), and lab notebooks. Manuscripts written with other authoring systems have been successfully ported to Manubot, including the [Bitcoin Whitepaper](#) [34] and [Project Rephetio manuscript](#) [35]. Finally, the Kipoi model zoo for genomics [36] uses Manubot's citation functionality to automatically extract model authors.

Citation utility

To make citation-by-identifier easily usable outside of Manubot manuscripts, we created the `manubot cite` command line utility, available as a [Python package](#). This utility takes a list of citations and returns either a rendered bibliography or CSL Data Items (i.e. JSON-formatted reference metadata). For example, the following command outputs a Markdown reference list for the two specified articles according to the bibliographic style of *PeerJ*:

```
manubot cite --render --format=markdown \  
  --csl=https://github.com/citation-style-language/styles/raw/master/peerj.csl \  
  pmid:29618526 doi:10.1038/550143a
```

Authorship

To determine authorship for the Deep Review, 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.

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. 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. The Deep Review author ordering procedure is not inherent to writing with Manubot but illustrates the authorship possibilities when all contributions are publicly tracked and recorded.

Discussion

Collaborative review manuscripts

The open scholarly writing Manubot enables has particular benefits for review articles, which present the state of the art in a scientific field [37]. Literature reviews are typically written in private by an invited team of colleagues. In contrast, broadly opening the process to anyone engaged in the topic — such that planning, organizing, writing, and editing occur collaboratively in a public forum where anyone is welcome to participate — can maximize a review’s value. Open drafting of reviews is especially helpful for capturing state-of-the-art knowledge about rapidly advancing research topics at the intersection of existing disciplines where contributors bring diverse opinions and expertise.

Writing review articles in a public forum allows review authors to engage with the original researchers to clarify their methods and results and present them accurately, as exemplified [here](#). Additionally, discussing manuscripts in the open generates valuable pre-publication peer review of preprints [16] or post-publication peer review [12,38,39]. Because incentives to provide public peer

review of existing literature [40] are lacking, open collaborative reviews — where authorship is open to anyone who makes a valid contribution — could help spur more post-publication peer review.

Additional collaborative writing projects

The Deep Review was not the first scholarly manuscript written online via an open collaborative process. In 2013, two dozen mathematicians created the 600-page Homotopy Type Theory book, writing collaboratively in LaTeX on GitHub [41,42]. Two technical books on cryptocurrency — [Mastering Bitcoin](#) and [Mastering Ethereum](#) — written on GitHub in AsciiDoc format have engaged hundreds of contributors. Both Homotopy Type Theory and Mastering Bitcoin continue to be maintained years after their initial publication. A 2017 perspective on the future of peer review was written collaboratively on Overleaf, with contributions from 32 authors [43]. While debate was raging over tightening the default threshold for statistical significance, nearly 150 scientists contributed to a Google Doc discussion that was condensed into a traditional journal commentary [44,45]. The greatest success to date of open collaborative writing is arguably Wikipedia, whose English version contains over 5.5 million articles. Wikipedia scaled encyclopedias [far beyond](#) any privately-written alternative. These examples illustrate how open collaborative writing can scale scholarly manuscripts where diverse opinion and expertise are paramount beyond what would otherwise be possible.

Open writing also presents new opportunities for distributing scholarly communication. Though it is still valuable to have versioned drafts of a manuscript with digital identifiers, journal publication may not be the terminal endpoint for collaborative manuscripts. After releasing the first version of the Deep Review [9], 14 new contributors updated the manuscript (Figure 2). Existing authors continue to discuss new literature, [creating a living document](#). Manubot provides an ideal platform for perpetual reviews [46,47].

Concepts for the future of scholarly publishing extend beyond collaborative writing [48,49]. Bookdown [50] and Pandoc Scholar [51] both extend traditional Markdown to better support publishing. Examples of continuous integration to automate manuscript generation include [gh-publisher](#) and Continuous Publishing [52], which was used to produce the book Opening Science [53]. Distill journal articles [54], Idyll [55], and Stencila [56] support manuscripts with interactive graphics and close integration with the underlying code. As an open source project, Manubot can be extended to adopt best practices from these other emerging platforms.

Several other open science efforts are GitHub-based like our collaborative writing process. ReScience [57], the Journal of Open Source Software [58], and some other [Open Journals](#) rely on GitHub for peer review and hosting. Distill uses GitHub for transparent peer review and post-publication peer review [59]. GitHub is increasingly used for resource curation [60], and collaborative scholarly reviews combine literature curation with discussion and interpretation.

Limitations

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.

In the Deep Review, we established [contributor guidelines](#) that discussed norms in the areas of text contribution, peer review, and authorship, which we identified in advance as potential areas of disagreement. Our contributor guidelines required verifiable participation: either directly attributable changes to the text or participation in the discussion on GitHub. These guidelines did not discuss broader community norms that may have improved inclusiveness. It is also important to consider how the move to an open contribution model affects under-represented minority members of the scientific community [15]. Recent work has identified clear social norms and processes as helpful to maintaining a collaborative culture [61]. Conferences and open source projects have used codes of conduct to establish these norms [62,63]. We would encourage the maintainers of similar projects to consider broader codes of conduct for project participants that build on social as well as academic norms.

Manubot in the context of open science

Science is undergoing a transition towards openness. The internet provides a global information commons, where scholarship can be publicly shared at a minimal cost. For example, open access publishing provides an economic model that encourages maximal dissemination and reuse of scholarly articles [64,65]. More broadly, open licensing solves legal barriers to content reuse, enabling any type of scholarly output to become part of the commons [66,67]. The opportunity to reuse data and code for new investigations, as well as a push for increased reproducibility, has begot a movement to make all research outputs public, unless there are bonafide privacy or security concerns [68,69,70]. New tools and services make it increasingly feasible to publicly share the unabridged methods of a study, especially for computational research, which consists solely of software and data.

Greater openness in both research methods and publishing creates an opportunity to redefine peer review and the role journals play in communicating science [43]. At the extreme is real-time open science, whereby studies are performed entirely in the open from their inception [71]. Many such

research projects have now been completed, benefiting from the associated early-stage peer review, additional opportunity for online collaboration, and increased visibility [35,72].

Manubot is an ideal authoring protocol for real-time open science, especially for projects that are already using an open source software workflow to manage their research. While Manubot does require technical expertise, the benefits are manifold. Specifically, Manubot demonstrates a system for publishing that is transparent, reproducible, immediate, permissionless, versioned, automated, collaborative, open, linked, provenanced, decentralized, hackable, interactive, annotated, and free of charge. These attributes empower integrating Manubot with an ecosystem of other community-driven tools to make science as open and collaborative as possible.

Acknowledgments

We would like to thank the authors of the Deep Review who helped us test collaborative writing with Manubot. The authors who responded favorably to being acknowledged are Paul-Michael Agapow, Amr M. Alexandari, Brett K. Beaulieu-Jones, Anne E. Carpenter, Travers Ching, Evan M. Cofer, Dave DeCaprio, Brian T. Do, Enrico Ferrero, David J. Harris, Michael M. Hoffman, Alexander A. Kalinin, Anshul Kundaje, Jack Lanchantin, Christopher A. Lavender, Benjamin J. Lengerich, Zhiyong Lu, Yifan Peng, Yanjun Qi, Gail L. Rosen, Avanti Shrikumar, Srinivas C. Turaga, Gregory P. Way, Laura K. Wiley, Stephen Woloszynek, Wei Xie, Jinbo Xu, and Michael Zietz. In addition, we thank Ogun Adebali, Evan M. Cofer, and Robert Gieseke for contributing to the Manubot template manuscript. We are grateful for additional Manubot discussion and testing by Alexander Dunkel, Ansel Halliburton, Achintya Rao, and other GitHub users.

Funding

DSH and CSG were supported by [Grant G-2018-11163](#) from the Alfred P. Sloan Foundation and [Grant GBMF4552](#) from the Gordon and Betty Moore Foundation. VSM was supported by [Grant RP150596](#) from the Cancer Prevention and Research Institute of Texas.

References

1. TechBlog: “Manubot” powers a crowdsourced “deep-learning” review

Jeffrey Perkel

Naturejobs (2018-02-20) <http://blogs.nature.com/naturejobs/2018/02/20/techblog-manubot-powers-a-crowdsourced-deep-learning-review/>

2. Crowdsourcing in biomedicine: challenges and opportunities

Ritu Khare, Benjamin M. Good, Robert Leaman, Andrew I. Su, Zhiyong Lu

Briefings in Bioinformatics (2016-01) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4719068/>

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, Michael Zietz, Michael M. Hoffman, ... Casey S. Greene

Journal of The Royal Society Interface (2018-04) <https://doi.org/10.1098/rsif.2017.0387>

4. Scientific writing: the online cooperative

Jeffrey M. Perkel

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

5. 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>

6. 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 Ternent, Stephen J. Eglen, Daniel S. Katz, ... Juan Antonio Vizcaíno

PLOS Computational Biology (2016-07-14) <https://doi.org/10.1371/journal.pcbi.1004947>

7. 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>

8. 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>

9. 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, Michael Zietz, Michael M Hoffman, ... Casey S. Greene

Cold Spring Harbor Laboratory (2017-05-28) <https://doi.org/10.1101/142760>

10. Reference Management

Martin Fenner, Kaja Scheliga, Sönke Bartling

Opening Science (2013-12-17) https://doi.org/10.1007/978-3-319-00026-8_8

11. Comparison of Select Reference Management Tools

Yingting Zhang

Medical Reference Services Quarterly (2012-01) <https://doi.org/10.1080/02763869.2012.641841>

12. Reviewing post-publication peer review.

Paul Knoepfler

Trends in genetics : TIG (2015-04-04) <https://www.ncbi.nlm.nih.gov/pubmed/25851694>

13. Decentralized Trusted Timestamping using the Crypto Currency Bitcoin

Bela Gipp, Norman Meuschke, André Gernandt

arXiv (2015-02-13) <https://arxiv.org/abs/1502.04015v1>

14. Twenty-Five Shades of Greycite: Semantics for referencing and preservation

Phillip Lord, Lindsay Marshall

arXiv (2013-04-26) <https://arxiv.org/abs/1304.7151v1>

15. Open science and open science

Laurent Gatto

(2017-06-05) <https://lgatto.github.io/open-and-open/>

16. Journal clubs in the time of preprints

Prachee Avasthi, Alice Soragni, Joshua N Bembenek

eLife (2018-06-11) <https://doi.org/10.7554/elife.38532>

17. On author versus numeric citation styles

Daniel Himmelstein

Satoshi Village (2018-03-12) <https://blog.dhimmel.com/citation-styles/>

18. TechBlog: Create the perfect bibliography with the CSL Editor

Jeffrey Perkel

Naturejobs (2017-05-03) <http://blogs.nature.com/naturejobs/2017/05/03/techblog-create-the-perfect-bibliography-with-the-csl-editor/>

19. JATS: Journal Article Tag Suite, version 1.1

National Information Standards Organization

(2015) <http://www.niso.org/standards/z39-96-2015/>

20. Journal Article Tag Suite 1.0: National Information Standards Organization standard of journal extensible markup language

Sun Huh

Science Editing (2014-08-18) <https://doi.org/10.6087/kcse.2014.1.99>

21. NISO Z39.96-201x, JATS: Journal Article Tag Suite

Mark H. Needleman

Serials Review (2012-09) <https://doi.org/10.1080/00987913.2012.10765464>

22. Collaborative software development made easy

Andrew Silver

Nature (2017-10-04) <https://doi.org/10.1038/550143a>

23. Reproducibility of computational workflows is automated using continuous analysis

Brett K Beaulieu-Jones, Casey S Greene

Nature Biotechnology (2017-03-13) <https://doi.org/10.1038/nbt.3780>

24. Developing a modern data workflow for evolving data

Glenda M Yenni, Erica M Christensen, Ellen K Bledsoe, Sarah R Supp, Renata M Diaz, Ethan P White, SK Morgan Ernest

Cold Spring Harbor Laboratory (2018-06-12) <https://doi.org/10.1101/344804>

25. Priority of discovery in the life sciences

Ronald D Vale, Anthony A Hyman

eLife (2016-06-16) <https://doi.org/10.7554/elife.16931>

26. The Grey Literature — Proof of prespecified endpoints in medical research with the bitcoin blockchain

Benjamin Gregory Carlisle

(2014-08-25) <https://www.bgcarlisle.com/blog/2014/08/25/proof-of-prespecified-endpoints-in-medical-research-with-the-bitcoin-blockchain/>

27. The most interesting case of scientific irreproducibility?

Daniel Himmelstein

Satoshi Village (2017-03-08) <http://blog.dhimmel.com/irreproducible-timestamps/>

28. Bitcoin: A Peer-to-Peer Electronic Cash System

Satoshi Nakamoto

(2017-09-20) <http://git.dhimmel.com/bitcoin-whitepaper/>

29. **OpenTimestamps: Scalable, Trustless, Distributed Timestamping with Bitcoin***Peter Todd*
(2018-02-10) <https://petertodd.org/2016/opentimestamps-announcement>

30. **eLife supports development of open technology stack for publishing reproducible manuscripts online**

Emily Packer

(2017-09-07) <https://elifesciences.org/for-the-press/e6038800/elifesciences-supports-development-of-open-technology-stack-for-publishing-reproducible-manuscripts-online>

31. **Sci-Hub provides access to nearly all scholarly literature**

Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene

eLife (2018-03-01) <https://doi.org/10.7554/elifesciences.32822>

32. **2017 in news: The science events that shaped the year**

Ewen Callaway, Davide Castelvecchi, David Cyranoski, Elizabeth Gibney, Heidi Ledford, Jane J. Lee, Lauren Morello, Nicky Phillips, Quirin Schiermeier, Jeff Tollefson, ... Alexandra Witze

Nature (2017-12-21) <https://doi.org/10.1038/d41586-017-08493-x>

33. **Vagelos Report Summer 2017**

Michael Zietz

Figshare (2017) <https://doi.org/10.6084/m9.figshare.5346577>

34. **How I used the Manubot to reproduce the Bitcoin Whitepaper**

Daniel Himmelstein

Steem (2017-09-20) <https://busy.org/@dhimmel/how-i-used-the-manubot-to-reproduce-the-bitcoin-whitepaper>

35. **Systematic integration of biomedical knowledge prioritizes drugs for repurposing**

Daniel Scott Himmelstein, Antoine Lizée, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, Sergio E Baranzini

eLife (2017-09-22) <https://doi.org/10.7554/elifesciences.26726>

36. **Kipoi: accelerating the community exchange and reuse of predictive models for genomics**

Ziga Avsec, Roman Kreuzhuber, Johnny Israeli, Nancy Xu, Jun Cheng, Avanti Shrikumar, Abhimanyu Banerjee, Daniel S Kim, Lara Urban, Anshul Kundaje, ... Julien Gagneur

Cold Spring Harbor Laboratory (2018-07-24) <https://doi.org/10.1101/375345>

37. **Ten Simple Rules for Writing a Literature Review**

Marco Pautasso

PLoS Computational Biology (2013-07-18) <https://doi.org/10.1371/journal.pcbi.1003149>

38. A Stronger Post-Publication Culture Is Needed for Better Science

Hilda Bastian

PLoS Medicine (2014-12-30) <https://doi.org/10.1371/journal.pmed.1001772>

39. Post-Publication Peer Review: Opening Up Scientific Conversation

Jane Hunter

Frontiers in Computational Neuroscience (2012) <https://doi.org/10.3389/fncom.2012.00063>

40. Post-publication peer review, in all its guises, is here to stay

Michael Markie

Insights the UKSG journal (2015-07-07) <https://doi.org/10.1629/uksg.245>

41. Homotopy Type Theory: Univalent Foundations of Mathematics

The Univalent Foundations Program

Institute for Advanced Study (2013) <https://homotopytypetheory.org/book/>

42. The HoTT book

Andrej Bauer

Mathematics and Computation (2013-06-20) <http://math.andrej.com/2013/06/20/the-hott-book/>

43. 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 (2017-11-01) <https://doi.org/10.12688/f1000research.12037.2>

44. Nearly 100 scientists spent 2 months on Google Docs to redefine the p-value. Here's what they came up with

Jop Vrieze

Science (2018-01-18) <https://doi.org/10.1126/science.aat0471>

45. Justify your alpha

Daniel Lakens, Federico G. Adolphi, Casper J. Albers, Farid Anvari, Matthew A. J. Apps, Shlomo E. Argamon, Thom Baguley, Raymond B. Becker, Stephen D. Benning, Daniel E. Bradford, ... Rolf A. Zwaan

Nature Human Behaviour (2018-02-26) <https://doi.org/10.1038/s41562-018-0311-x>

46. A proposal for regularly updated review/survey articles: "Perpetual Reviews"

David L. Mobley, Daniel M. Zuckerman

arXiv (2015-02-03) <https://arxiv.org/abs/1502.01329v2>

47. Why we need the Living Journal of Computational Molecular Science

David L. Mobley, Michael R. Shirts, Daniel M. Zuckerman

Living Journal of Computational Molecular Science (2017-08-22) <http://www.livecomsjournal.org/article/2031-why-we-need-the-living-journal-of-computational-molecular-science>

48. The arXiv of the future will not look like the arXiv

Alberto Pepe, Matteo Cantiello, Josh Nicholson

Authorea <https://doi.org/10.22541/au.149693987.70506124>

49. TechBlog: C. Titus Brown: Predicting the paper of the future

C. Titus Brown

Naturejobs (2017-06-01) <http://blogs.nature.com/naturejobs/2017/06/01/techblog-c-titus-brown-predicting-the-paper-of-the-future/>

50. bookdown

Yihui Xie

Chapman & Hall/CRC The R Series (2016-12-21) <https://doi.org/10.1201/9781315204963>

51. 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>

52. Continuous Publishing

Martin Fenner

Gobbledygook (2014-03-10) <http://blog.martinfenner.org/2014/03/10/continuous-publishing/>

53. Opening Science *Springer International Publishing* (2014) <https://doi.org/10.1007/978-3-319-00026-8>

54. The Building Blocks of Interpretability

Chris Olah, Arvind Satyanarayan, Ian Johnson, Shan Carter, Ludwig Schubert, Katherine Ye, Alexander Mordvintsev

Distill (2018-03-06) <https://doi.org/10.23915/distill.00010>

55. Announcing idyll.pub

Matthew Conlen, Andrew Osheroff

Idyll (2018-06-26) <https://idyll.pub/post/announcing-idyll-pub-0a3eff0661df3446a915700d/>

56. Stencila – an office suite for reproducible research

Michael Aufreiter, Aleksandra Pawlik, Nokome Bentley

eLife Labs (2018-07-02) <https://elifesciences.org/labs/c496b8bb/stencila-an-office-suite-for-reproducible-research>

57. Sustainable computational science: the ReScience initiative

Nicolas P. Rougier, Konrad Hinsén, Frédéric Alexandre, Thomas Arildsen, Lorena A. Barba, Fabien

C.Y. Benureau, C. Titus Brown, Pierre de Buyl, Ozan Caglayan, Andrew P. Davison, ... Tiziano Zito
PeerJ Computer Science (2017-12-18) <https://doi.org/10.7717/peerj-cs.142>

58. 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
PeerJ Computer Science (2018-02-12) <https://doi.org/10.7717/peerj-cs.147>

59. Distill Update 2018

Distill Editors
Distill (2018-08-14) <https://doi.org/10.23915/distill.00013>

60. The appropriation of GitHub for curation

Yu Wu, Na Wang, Jessica Kropczynski, John M. Carroll
PeerJ Computer Science (2017-10-09) <https://doi.org/10.7717/peerj-cs.134>

61. Innovating Collaborative Content Creation: The Role of Altruism and Wiki Technology

Christian Wagner, Pattarawan Prasarnphanich
2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07) (2007) <https://doi.org/10.1109/hicss.2007.277>

62. Code of conduct evaluations *Geek Feminism Wiki* (2017-06-13) http://geekfeminism.wikia.com/wiki/Code_of_conduct_evaluations?oldid

63. Contributor Covenant: A Code of Conduct for Open Source Projects

Coraline Ada Ehmke
(2014) <https://www.contributor-covenant.org/>

64. The academic, economic and societal impacts of Open Access: an evidence-based review

Jonathan P. Tennant, François Waldner, Damien C. Jacques, Paola Masuzzo, Lauren B. Collister, Chris. H. J. Hartgerink
F1000Research (2016-09-21) <https://doi.org/10.12688/f1000research.8460.3>

65. 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>

66. The Legal Framework for Reproducible Scientific Research: Licensing and Copyright

Victoria Stodden
Computing in Science & Engineering (2009-01) <https://doi.org/10.1109/mcse.2009.19>

67. Legal confusion threatens to slow data science

Simon Oxenham

Nature (2016-08-03) <https://doi.org/10.1038/536016a>

68. Enhancing reproducibility for computational methods

V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. A. Ioannidis, M. Taufer

Science (2016-12-08) <https://doi.org/10.1126/science.aah6168>

69. The case for open computer programs

Darrel C. Ince, Leslie Hatton, John Graham-Cumming

Nature (2012-02-22) <https://doi.org/10.1038/nature10836>

70. 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>

71. This revolution will be digitized: online tools for radical collaboration

C. Patil, V. Siegel

Disease Models & Mechanisms (2009-04-30) <https://doi.org/10.1242/dmm.003285>

72. Publishing the research process

Daniel Mietchen, Ross Mounce, Lyubomir Penev

Research Ideas and Outcomes (2015-12-17) <https://doi.org/10.3897/rio.1.e7547>