

# Open collaborative writing with Manubot

This manuscript ([permalink](#)) was automatically generated from [greenelab/meta-review@799d876](#) on July 21, 2018.

## Authors

---

- **Daniel S. Himmelstein**

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

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

- **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 · Funded by CPRIT RP150596

- **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; Morgridge Institute for Research · Funded by NIH U54AI117924

# Abstract

---

Scientific literature reviews are ideal for open, collaborative writing. Allowing any interested individual to contribute as an author can strengthen a review, providing broad and fair coverage of the subject matter. However, the traditional multi-author writing process breaks down at scale. We present techniques for overcoming the challenges of open manuscript writing. These include approaches for managing distributed authors and our new software, named Manubot, for automating citation and manuscript building.

# Introduction

---

Openness in research benefits the researchers who practice it [1], their scientific peers, and the public. Improving the availability of code and data helps ensure the methods are transparent and the results can be reproduced [2]. Publishing in an open access journal increases the number of readers and reviewers [3]. Open scholarly writing, a form of crowdsourcing [4], has particular benefits for review articles, which present the state of the art in a scientific field [5]. Literature reviews are typically written in private by a closed 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.

Based on our experience leading a recent open review [6], we discuss the pros and cons of open collaborative reviews as well as how they contribute to a living and frequently updated literature. Our review manuscript [7], code-named the Deep Review, surveyed deep learning's role in biology and precision medicine, a research area undergoing explosive growth. In addition, we introduce [Manubot](#), the infrastructure we created to enable open manuscript writing online for Deep Review, which was subsequently adopted by other projects.

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's adaptable for collaborative writing. From the start, we made the GitHub repository public, applying a [Creative Commons Attribution License](#) to the manuscript. Next, 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.

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 post-publication peer review [10,8,9] or pre-publication peer review of preprints [11]. Because incentives are commonly lacking to provide public peer review of existing literature [12], open collaborative reviews — where authorship is open to anyone who makes a valid contribution — could help spur more post-publication peer review. 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.

To address these challenges, we developed a manuscript writing process using the Markdown language, the GitHub platform, and our new Manubot tool for automating manuscript generation.

## 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 [13] 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, 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.<sup>1</sup> Additional functionality, such as bibliography management, is available by editing the Word document stored in OneDrive with the paid Word desktop application.<sup>2</sup> Conversations about modifications take place on the document as comments, annotations, or unsaved chats. There is no separate integrated forum for discussing and editing revisions.<sup>3</sup> Simultaneous edits online can be attributed to a single author in a git commit.

<b>Feature</b>	<b>Manubot</b>	<b>Authorea + BibTeX</b>	<b>Overleaf + BibTeX</b>	<b>Google Docs + Paperpile</b>	<b>Word Online<sup>1</sup></b>	<b>Markdown on GitHub</b>
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 <sup>2</sup>	No <sup>2</sup>	No <sup>2</sup>	No <sup>2</sup>	Yes
Character-level provenance for text	Yes	No (versions tracked by day)	No <sup>3</sup>	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 + BibTeX	Google Docs + Paperpile	Word Online <sup>1</sup>	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
Technial expertise required	Yes	Yes	Yes	No	No	Yes

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 [14,15] 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 16. 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 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. The process of writing and revising material can be orchestrated through GitHub with a web browser.

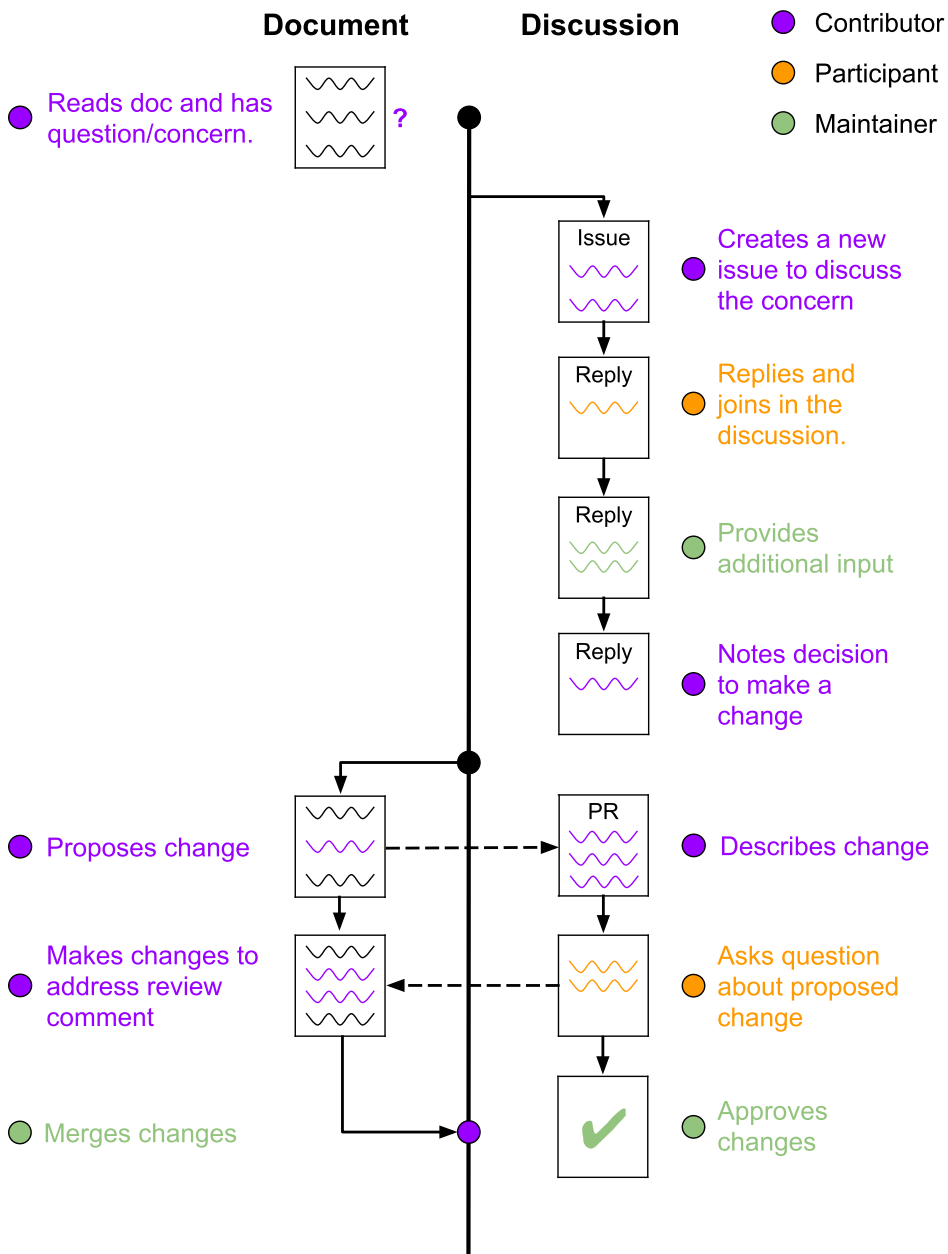


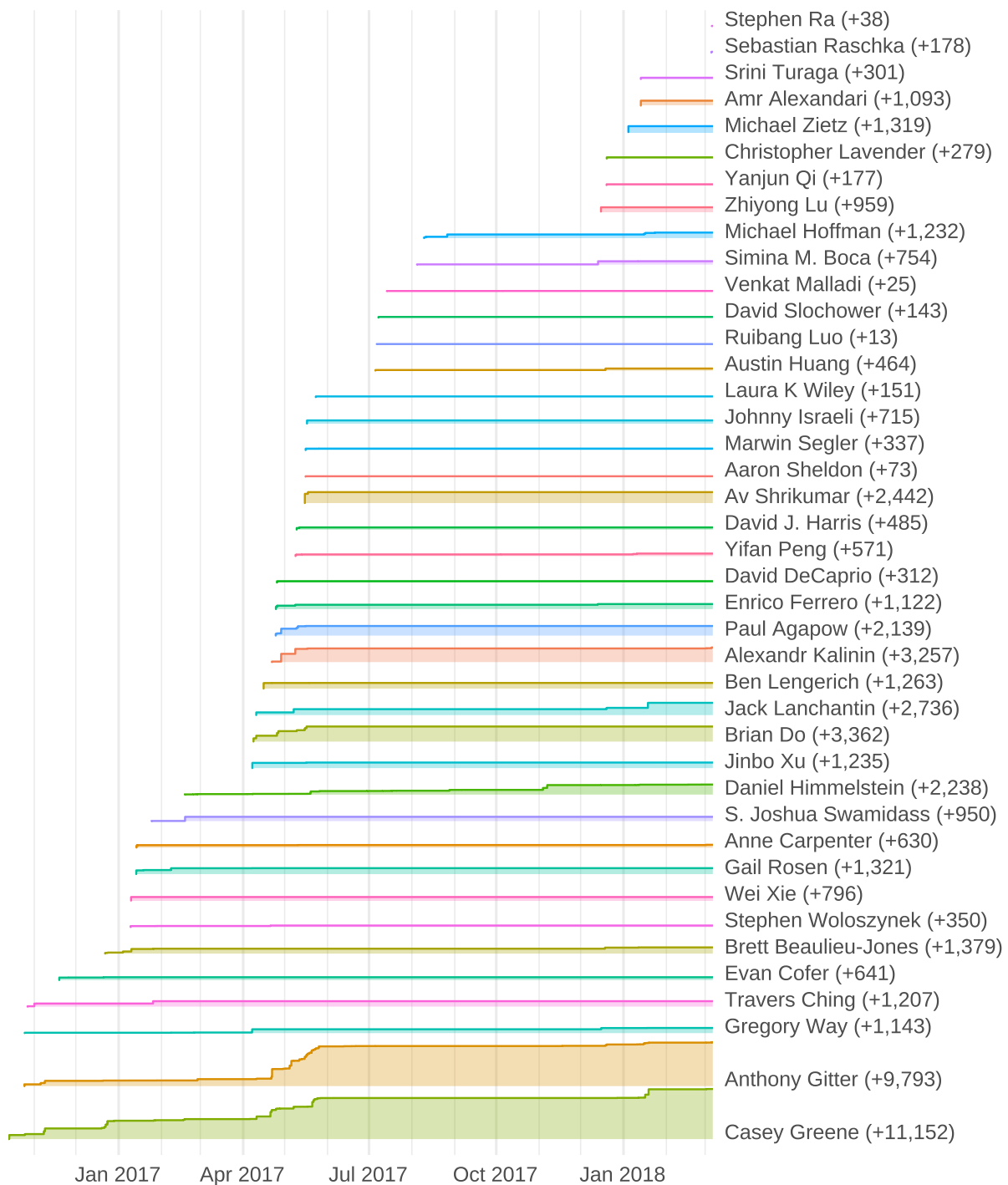
Figure 1: **Deep Review editing workflow.** Any reader can become a contributor by proposing a change through a pull request. In this example, the contributor opens an issue to discuss a manuscript modification. A maintainer and additional participant provide feedback, and the maintainer recommends creating a pull request to update the text. The contributor creates the pull request, it is reviewed by a maintainer and a participant, and the contributor updates the pull request in response. Once the pull request is approved, the maintainer merges the changes into the official version of the manuscript.

The Deep Review [issue](#) and [pull request](#) on protein-protein interactions demonstrate this process in practice. A new contributor identified a relevant research topic that was missing from the review manuscript with examples of how the literature would be synthesized. A maintainer confirmed that this was a desirable topic and referred to related open 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 [17]. 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 our 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. Deep Review was initiated in August 2016, and the first complete manuscript was released as a preprint [18] 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 [7]. As of June 15, 2018, the Deep Review repository accumulated 755 git commits, 315 merged pull requests, 537 issues, and 616 GitHub stars.

## Manubot

---

We developed Manubot, a system for writing scholarly manuscripts via GitHub. 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 Items](#), an open standard that is widely supported by reference managers [19,20]. 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 `Avasthi2018_preprints` represents the DOI identifier `doi:10.7554/eLife.38532`.

Identifier	Metadata source	Example citation	Processed citation
Digital Object Identifier (DOI)	DOI <a href="#">Content Negotiation</a>	<code>doi:10.1098/rsif.2017.0387</code>	[7]
PubMed Identifier (PMID)	NCBI's <a href="#">E-utilities</a>	<code>pmid:25851694</code>	[8]
PubMed Central Identifier (PMCID)	NCBI's <a href="#">Citation Exporter</a>	<code>pmcid:PMC4719068</code>	[4]
arXiv identifier	<a href="#">arXiv API</a>	<code>arxiv:1502.04015v1</code>	[21]
URL	<a href="#">Greycite</a> [22]	<code>url:https://lgatto.github.io/open-and-open/</code>	[23]
Tag	Source for tagged identifier	<code>tag:Avasthi2018_preprints</code>	[11]

Manubot formats bibliographies according to a [CSL](#) specification. As a result, users can choose from thousands of existing CSL styles or use Manubot's default style. Styles define how references are constructed from bibliographic metadata, controlling layout details such as the maximum number of authors to list per reference. Thousands of journals have [predefined styles](#). 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 [24,25,26]. Going forward, we hope to integrate Manubot with the larger JATS ecosystem. For now however, 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) [27,28,29], specifically [Travis CI](#) at the moment, to monitor for 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. Since 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.

For this article, the source GitHub repository 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/3ec66cd440e78dac33f4700f1cecbc4d95b20139/> (`(TODO: update commit)`), 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 [30]. 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. We implemented support for [OpenTimestamps](#) in Manubot 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 [21,31,32,33]. For Manubot manuscripts, scientific precedence can now be indisputably established, and timestamps protect against attempts to rewrite a manuscript's history. Such timestamping practices help ensure accurate histories, potentially alleviating certain authorship or priority disputes. Since all bitcoin transactions are competing 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 [34], but there can be a lag of a few hours before the transaction is made. We judged this to be suitable for the purposes of scientific writing.

We designed Manubot to power the next generation of scholarly manuscript. Manubot transforms publication, making it permissionless, reproducible, free of charge, and largely open source. Manubot 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. One direction Manubot is working towards is end-to-end document reproducibility, where every figure or piece of data in a manuscript can be traced back to its origin [35]. Already, Manubot 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 get 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 [36]. 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. A larger [set of revisions](#) that was merged a month later 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. Of interest, the Sci-Hub Coverage Study preprint was the [most viewed](#) 2017 *PeerJ Preprint*, while the Deep Review was the most viewed 2017 *bioRxiv* preprint [37]. 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, including in 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](#) [38], [literature reviews](#), and lab notebooks. Finally, manuscripts written with other authoring systems have been successfully ported to Manubot, including the [Bitcoin Whitepaper](#) [39] and [Project Rephetio manuscript](#) [40].

## 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 [7], 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.

## Discussion

---

The Deep Review was not the first scholarly manuscript written online via an open collaborative process. Interestingly, the collaborative scholarly works (of which we’re aware) tend to have dealt with rapidly developing topics. In 2013, two dozen mathematicians created the 600-page “HoTT Book” on “Homotopy Type Theory”, 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 The HoTT Book 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. As the examples above illustrate, open collaborative writing appears to scale scholarly manuscripts where diverse opinion and expertise are paramount beyond what would otherwise be possible.

Concepts for the future of scholarly publishing extend beyond collaborative writing [46,47]. Bookdown [48] and Pandoc Scholar [49] both extend traditional Markdown to better support publishing. Examples of continuous integration to automate manuscript generation include [gh-publisher](#) and Continuous Publishing [50], which was used to produce the book Opening Science [51]. Distill journal articles [52], Idyll [53], and Stencila [54] 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.

Many others have embraced open science principles and piloted open approaches toward drug discovery [40,55], data management [56,57,58], and manuscript review [59]. Several of these open science efforts are GitHub-based like our collaborative writing process. The ReScience [60], the Journal of Open Source Software [61], and some other [Open Journals](#) rely on GitHub for peer

review and hosting. GitHub is also increasingly used for resource curation [62], and collaborative scholarly reviews combine literature curation with discussion and interpretation.

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 our open review paper, 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. We maintained our guidelines, even when a case arose where two authors had written text together but only one had directly attributable changes and participation. 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 [23]. Recent work has identified clear social norms and processes as helpful to maintaining a collaborative culture [63]. Conferences and open source projects have used codes of conduct to establish these norms [64,65]. We would encourage the maintainers of similar projects to consider broader codes of conduct for project participants that establish on social as well as academic norms.

Open writing presents new opportunities for scholarly communication. 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 [18], 14 new contributors have updated the manuscript (Figure 2) and existing authors continue to discuss new literature, [creating a living document](#). Manubot provides an ideal platform for perpetual reviews [66,67].

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

---

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, Alexandr 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.



# 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. A manifesto for reproducible science

Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, John P. A. Ioannidis  
*Nature Human Behaviour* (2017-01-10) <https://doi.org/10.1038/s41562-016-0021>

## 3. Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar, Roger S. Day, Douglas B. Fridsma  
*PLoS ONE* (2007-03-21) <https://doi.org/10.1371/journal.pone.0000308>

## 4. 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/>

## 5. Ten Simple Rules for Writing a Literature Review

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

## 6. 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/>

## 7. 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>

## 8. Reviewing post-publication peer review.

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

**9. 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>

**10. Post-Publication Peer Review: Opening Up Scientific Conversation**

Jane Hunter

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

**11. 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>

**12. 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>

**13. Scientific writing: the online cooperative**

Jeffrey M. Perkel

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

**14. 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>

**15. 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>

**16. 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>

**17. 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>

**18. 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>

#### 19. Reference Management

Martin Fenner, Kaja Scheliga, Sönke Bartling

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

#### 20. Comparison of Select Reference Management Tools

Yingting Zhang

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

#### 21. Decentralized Trusted Timestamping using the Crypto Currency Bitcoin

Bela Gipp, Norman Meuschke, André Gernandt

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

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

Phillip Lord, Lindsay Marshall

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

#### 23. Open science and open science

Laurent Gatto

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

#### 24. JATS: Journal Article Tag Suite, version 1.1

National Information Standards Organization

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

#### 25. 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>

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

Mark H. Needleman

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

#### 27. Collaborative software development made easy

Andrew Silver

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

#### 28. 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>

**29. Developing a modern data workflow for living 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>

**30. Priority of discovery in the life sciences**

Ronald D Vale, Anthony A Hyman

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

**31. 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/>

**32. The most interesting case of scientific irreproducibility?**

Daniel Himmelstein

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

**33. Bitcoin: A Peer-to-Peer Electronic Cash System**

Satoshi Nakamoto

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

**34. OpenTimestamps: Scalable, Trustless, Distributed Timestamping with Bitcoin***Peter Todd*

(2018-02-10) <https://petertodd.org/2016/opentimestamps-announcement>

**35. 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>

**36. 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/elife.32822>

**37. 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>

**38. Vagelos Report Summer 2017**

Michael Zietz

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

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

Daniel Himmelstein

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

**40. 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/elife.26726>

**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. 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>

**47. 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/>

**48. bookdown**

Yihui Xie

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

**49. 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>

**50. Continuous Publishing**

Martin Fenner

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

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

**52. 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>

**53. Announcing idyll.pub**

Matthew Conlen, Andrew Osheroff

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

**54. 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>

**55. An open source pharma roadmap**

Manica Balasegaram, Peter Kolb, John McKew, Jaykumar Menon, Piero Olliaro, Tomasz Sablinski, Zakir Thomas, Matthew H. Todd, Els Torreele, John Wilbanks

*PLOS Medicine* (2017-04-18) <https://doi.org/10.1371/journal.pmed.1002276>

**56. 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>

**57. 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>

**58. First, design for data sharing**

John Wilbanks, Stephen H Friend

*Nature Biotechnology* (2016-04) <https://doi.org/10.1038/nbt.3516>

**59. 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>

**60. 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 Buyt, Ozan Caglayan, Andrew P. Davison, ... Tiziano Zito

*PeerJ Computer Science* (2017-12-18) <https://doi.org/10.7717/peerj-cs.142>

**61. 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>

**62. 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>

**63. 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>

**64. Code of conduct evaluations** *Geek Feminism Wiki* (2017-06-13) [http://geekfeminism.wikia.com/wiki/Code\\_of\\_conduct\\_evaluations?oldid](http://geekfeminism.wikia.com/wiki/Code_of_conduct_evaluations?oldid)

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

Coraline Ada Ehmke

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

**66. 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>

**67. 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>