Manubot 2018 development proposal

This manuscript was automatically generated from [greenelab/manufund-2018@5d08b30](#) on December 22, 2017.

Authors

- **Daniel S. Himmelstein**

  ⓘ 0000-0002-3012-7446 · ⓞ dhimmel · 🐦 dhimmel

  Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania

- **Casey S. Greene**

  ⓘ 0000-0001-8713-9213 · ⓞ cgreene · 🐦 GreeneScientist

  Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania

Proposal

Manubot is a system for writing scholarly documents on GitHub. Manubot aims to **transform publishing** to be transparent & reproducible, immediate & permissionless, versioned & automated, collaborative & open, linked & provenanced, decentralized & hackable, interactive & annotated, and free of charge.

Manubot transfers the lessons from the **open source software movement** to academic writing. Documents are written in the open, creating an easy path for readers to become contributors. We recommend following a workflow where anyone can propose changes, which are then reviewed by project members and revised as necessary before being incorporated (for example, see how this proposal was created).

Manubot is based on a collection of open source **tools and standards**. Manuscripts are written in markdown, using git for version control and GitHub for collaboration and review. Users can cite standard persistent identifiers and Manubot retrieves bibliographic details to automatically create a reference list. Conversion between formats (e.g. from markdown to HTML, PDF, and DOCX) is done using Pandoc. When the manuscript changes, continuous integration automatically rebuilds and deploys it [1,2].

Currently, the Manubot consists of two repositories: `greenelab/manubot-rootstock` (a blank manuscript with the required files) and `greenelab/manubot` (a backend Python package). As of December 2017, 10 individuals have contributed to the **Manubot codebase**, which has 33 ⭐s on GitHub. For more information on the Manubot, see these slides as well as the in-progress Meta Review.

We initially created the Manubot to enable the **Deep Review**, a review paper on deep learning that we collaboratively drafted on GitHub. Over 30 authors contributed to the Deep Review [3], using issues to discuss source material and pull requests to propose manuscript changes. The community appreciated the breadth of perspectives on a quickly-developing topic and the Deep Review was the most-viewed *bioRxiv* preprint from April–June 2017 [4]. We've also used the Manubot to write our Sci-Hub Coverage Study — the second most viewed *PeerJ Preprint* ever — and to reproduce the Bitcoin Whitepaper [5]. In the Manubot's first year, its collaborative, open, and review-driven *authoring process* yielded two of the most popular preprints of 2017.

In 2018, we seek to expand Manubot's capabilities by accomplishing four feature-driven specific aims and one outreach aim:

1. JATS is the predominant archival format for scholarly manuscripts that is optimized for machine-readability and interoperability. We will implement JATS export and incorporate an existing **JATS viewer**, such as Lens, into Manubot webpages.

2. Currently all changes to a manuscript's source are tracked using git. We will implement **rendered diff functionality** to show how a manuscript changed from the reader's perspective and make creating "tracked changes" documents for journal resubmission easy.

3. **End-to-end reproducibility** is when the provenance of every manuscript result can be traced to its source. We will continue to evaluate and improve the Manubot template variables functionality, which allows users to embed results and method parameters directly from the analyses that created them.

4. The Manubot's **citation metadata** system is a powerful and versatile approach for bibliographic information. We will continue to expand Manubot's citation-by-persistent-ID capabilities.

5. Currently Manubot requires considerable technical expertise to setup and some technical expertise to use. The current userbase consists of researchers or publishers who are using tools like git and continuous integration for other tasks, so we will **improve documentation, develop user stories, and perform outreach activities** at relevant conferences.

Our **long-term vision** is an improved ecosystem of scholarly communication. Imagine you could click an author to highlight only the text they wrote. Or highlight all results and figures that descend from source code contributions by the selected author. Or highlight all results that depend on a selected upstream experiment or database. Or, if a new version of software used in the manuscript or an updated dataset is released, rebuild the manuscript with the updated version and see the resulting changes… regardless of whether you're the original authors or a curious third-party! Manubot forms the foundation for the traceability of all aspects of a scholarly manuscript, which is the key principle that underlies this potential future.

Now is the time to bring the innovation and ethos of open source software development to the scholarly manuscript. Join us in powering the next generation of scholarly manuscript.

References

1. **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

2. **Collaborative software development made easy**
Andrew Silver
Nature (2017-10-04) https://doi.org/10.1038/550143a

3. **Opportunities And Obstacles For Deep Learning In Biology And Medicine**
Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Wei Xie, Gail L. Rosen, … Casey S. Greene
Cold Spring Harbor Laboratory (2017-05-28) https://doi.org/10.1101/142760

4. **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 (2017-12-18) https://doi.org/10.1038/d41586-017-08493-x

5. **How I used the Manubot to reproduce the Bitcoin Whitepaper**
Daniel Himmelstein
Steemit (2017-09-20) https://steemit.com/manubot/@dhimmel/how-i-used-the-manubot-to-reproduce-the-bitcoin-whitepaper