

Final Report for Grant G-2018-11163

Grant Details

Principal Investigator: Daniel Himmelstein

Institution: University of Pennsylvania

Grant Title: To support continued development of Manubot, a git-native authoring tool for scientific manuscripts

Grant Webpage: <https://sloan.org/grant-detail/8501>

Grant amount: \$48,784

Grant identifier: G-2018-11163

Grant period: 2018-09-01 to 2019-08-31

We'd like to thank the Sloan Foundation for their support of the Manubot project. It has been an exciting year as the project has matured from a tool that worked for a single collaborative manuscript to a general purpose tool. We are delighted to have a diverse variety of manuscripts written with Manubot ranging from a [French article on visual illusions](#) to a [software paper on the thermodynamics of force fields](#).

We continue to strive towards a future where authoring and publishing are transparent, automated, and free of charge. Where scholars from around the world can collaborate openly to produce rigorous and reproducible research. And finally where outdated aspects of publishing are replaced by community-driven, open source workflows. Manubot hopes to lead the way in demonstrating what the future of publishing will look like!

Please find our report on the specific outcome metrics below.

Outcome metric: Launch of new Manubot features

- We created more versatile and robust infrastructure for **citation-by-persistent-identifier**. We added support for citing [ISBNs](#), [Wikidata IDs](#), and [shortDOIs](#). In addition, Manubot [now](#) allows “raw” citations for works without any persistent identifier, where metadata must be provided by the user. Manubot now retrieves PubMed Central metadata [from](#) the new NCBI Literature Citation Exporter as well as URL metadata [from](#) Zotero's translation-server, which provides custom metadata extractors for over 500 different resources.

Finally, we noticed some metadata services would return improperly formatted CSL JSON that would crash pandoc-citeproc (the program Manubot users to render bibliographies). Therefore, we [now compare](#) CSL JSON to its official schema and remove any invalid elements until it validates.

- We hired a **front-end** software engineer, [Vince Rubinitti](#), who started on 2019-01-07 and was partially supported by this grant. Vince's expertise presented an opportunity for us to overhaul the appearance, interactivity, and accessibility of rendered Manubot manuscripts. Vince created a **new theme and suite of plugins** for the HTML output, which provide a [state-of-the-art interface](#) for reading scholarly manuscripts. The CSS and Javascript source files providing this interface were designed to work with other Pandoc HTML outputs, such that we expect adoption beyond just Manubot manuscripts.
- Manubot is perhaps the only tool for writing scholarly manuscripts that uses **timestamping** to produce irrefutable proofs of manuscript existence at a given point in time. Timestamps can be used to detect when a historical record has been tampered with, thereby helping to ensure academic integrity and avoid authorship disputes. Manubot's continuous integration pipeline [now upgrades](#) timestamps after they are encoded into the Bitcoin blockchain, such that users get the full benefits of timestamps with no cost or inconvenience. Manubot's usage of timestamps was mentioned in a *Nature News* [toolbox article](#) from 2019-02-04.
- Previously, the effect of pull requests (proposals to modify a manuscript) on the rendered manuscript was not available online. This was a major [pain point for users](#). Project co-lead Anthony Gitter implemented a solution using AppVeyor continuous integration, which supports "build artifacts". Now [officially part of](#) Rootstock (Manubot's template repository), the @AppVeyorBot comments on pull requests with a download link for the **rendered PDF manuscript**. [Example](#):



- The ability to **export manuscripts to wide-variety of formats**, maximizing interoperability while avoiding lock-in, has always been a major attraction of Manubot. While Pandoc supports conversions between most formats, high fidelity conversion is challenging. While we would like many of these problems to be solved upstream, we continue to optimize certain common conversions. We now offer [athenapdf](#) for creating PDFs from HTML manuscript, which performs similarly to printing to PDF in a modern browser. Development to support [JATS](#) and [LaTeX](#) output is still ongoing.
- One development goal has been **rendered diffs** between manubot versions, similar to track changes in a word processor. We have been [exploring](#) various solutions and have

[prototyped](#) several options. None of the options has been reliable enough for production, and we continue to monitor the space.

- **Setting up new manuscripts** has always been an obstacle to user adoption, since it is more technically challenging than contributing to an existing manuscript. We improved our [setup documentation](#). The biggest challenge for users was configuring continuous integration with the required secrets to deploy rendered manuscripts back to GitHub. We switched to an easier way of providing secrets to continuous integration. This removed all remaining barriers to setting up manuscripts on Windows. Although seemingly a [small change](#), considerable testing and user observation was required to ensure the approach was well documented and worked across platforms.
- We have made **many improvements to the Manubot software** across the board. Michael Hoffman provided [advice](#) (and pull requests) on how to harden our shell scripts. On Travis CI, we [improved configuration](#) of the deploy stage. We [modified](#) our conda environment to improve reliability and persistence. We refactored the [manubot/manubot](#) Python package, including adopting lazy imports, clear submodule separation, and improved logging and error messages. We're also [working](#) with other experts in the space to develop and adopt a more standardized terminology for citation related concepts.
- We have continued to improve Manubot's [bibliographic style](#) for references, including how to address corner cases. The style can be used anywhere CSL (Citation Style Language) is supported and [provides](#) a more modern & informative layout compared to traditional references. [Example](#):

🚩 **2. Open Science by Design: Realizing a Vision for 21st Century Research**
National Academies of Sciences, Engineering, and Medicine
National Academies Press (2018-08-09) <https://doi.org/gfxzcc4>
DOI: [10.17226/25116](https://doi.org/10.17226/25116) · PMID: [30212065](https://pubmed.ncbi.nlm.nih.gov/30212065/) · ISBN: [9780309476249](https://www.isbn-international.org/product/9780309476249)

Outcome metric: Growth of Manubot community

- We created a **project homepage** at <https://manubot.org/>, which is a more welcoming introduction for most users than a GitHub readme.
- The website contains a [catalog of manuscripts](#) that were made with Manubot. Manuscripts are sorted by most recently updated and include links to the corresponding preprint and journal versions when available. Manuscripts can be added to the catalog by making a pull request to the [manubot/catalog](#) repository. The catalog is then automatically generated and updated. Currently, the catalog contains

28 manuscripts, since it mostly consists of manuscripts that have at least reached preprint stage. We are considering how to expand it and encourage more community curation.

- We wrote a **software paper** describing Manubot which was published in *PLOS Computational Biology* on 2019-06-24 ([manubot version](#), [journal version](#)). While the journal does not publish peer review reports, since our authoring process was entirely public, the peer reviews and our responses [are available](#) in the manuscript repository. Furthermore, we are continuing to make some updates to the manuscript post publication, which we alert readers to by writing: “This version of the manuscript [contains changes](#) subsequent to the journal publication.”
- During the grant period, there were 5 **code contributors** to [manubot/rootstock](#), 4 to [manubot/manubot](#), and 4 to [manubot/catalog](#). Several of these contributors were first-time contributors (i.e. the first commit from the account was during the grant period) including [@vincerubineti](#), [@michaelmhoffman](#), [@nichtich](#), [@trang1618](#), and [@danich1](#). Many more individuals got involved by opening GitHub issues. Together over 150 issues have been opened on the [manubot/manubot](#) and [manubot/rootstock](#) repositories. We encourage users to open issues liberally, not just for bug reports and feature requests, but also for help requests and discussion of related projects.
- We organized **outreach events** including workshops and presentations on Manubot. At FORCE2018 in Montreal, Daniel Himmelstein gave a presentation titled “Revolutionizing scholarly writing with Manubot” ([recording](#), [slides](#)). At the Pacific Symposium on Biocomputing 2019, we led a workshop where 17 participants collaboratively authored a [manuscript titled](#) “The people of the 24th annual Pacific Symposium on Biocomputing in Hawaii.” At the Researcher2Reader 2019 conference in London, Daniel Himmelstein [led a workshop](#) on citation-by-persistent-identifier, which sought to introduce publishers to the topic and collect their feedback. Daniel also discussed Manubot in [two lectures](#) at the University of Utah, one at the cancer center and another to students of the [DeCART](#) Summer Program. Anthony Gitter presented Manubot [as part of eLife Community Call](#) on 2018-06-28.
- We created a Manubot manuscript called [try-manubot](#), where users can make any contribution to try out authoring with Manubot. This manuscript provides a quick way for users to evaluate Manubot without setting up a manuscript of their own. Project member David Slochower produced a [getting started video](#) demonstrating how to propose a change to a manuscript entirely through the GitHub web interface.
- We obtained the [@manubot username](#) on GitHub, which had previously belonged to an inactive user and migrated repositories from the [@greenelab](#) organization.

- We created a Manubot [logo](#) and printed hex stickers with the logo that we are distributing and [mailing](#) to interested users around the world.
- We **improved documentation**, though the website and the [usage](#) and delete-me files in Rootstock. The delete-me shows the formatting options available for Manubot manuscript ([raw](#), [rendered](#)).
- We adopted **semantic versioning** for the manubot/manubot python package and [publish releases](#) with detailed changelogs. The package is available [from PyPI](#) (the Python Packaging Index).
- We created the “**manubot cite**” command line utility, that allows generating references for persistent identifiers from a terminal. We produced [terminal recordings](#) demonstrating the functionality. This utility is [now gaining](#) adoption outside of Manubot manuscript users.
- A researcher from Argentina, Humberto Debat, contacted us with interest in translating our Sci-Hub Coverage Study written with Manubot and published in *eLife* in 2018. We helped Humberto create the **Spanish translation** in Manubot, now available at <https://greenelab.github.io/scihub-manuscript-es/>.

Estimaciones recientes sugieren que los muros de pago en la web limitan el acceso a tres cuartos de la literatura académica [2,3,4]. El movimiento de acceso abierto persevera por remediar esta situación [5]. Después de décadas de esfuerzo por parte de la comunidad de acceso abierto [6], casi el 50% de los artículos recientemente publicados están disponibles sin muros de pagos [2,7,8].

▼ Inglés original

Recent estimates suggest paywalls on the web limit access to three-quarters of scholarly literature [2,3,4]. The open access movement strives to remedy this situation [5]. After decades of effort by the open access community [6], nearly 50% of newly published articles are available without paywalls [2,7,8].

- Manubot relies on an **ecosystem of open source** software and provided **frequent contributions to related projects**.

We began using Zotero’s [translation-server](#) to generate bibliographic metadata for several types of citations. While Zotero makes the source code available, they do not host a public instance. We have [set up](#) a public instance at the URL translate.manubot.org, which is used by not only by Manubot but also now by unaffiliated projects [such as](#) [citation.js](#). We have contributed several suggestions and/or proposed changed to [translation-server](#), as well as the upstream [translators](#) and [zotero](#) repositories.

Citation Styles Language has become the primary alternative to LaTeX/BibTeX for specifying bibliographic styles. Manubot, like most other open tools in scholarly publishes, uses CSL to generate bibliographies. We [have made](#) substantial contributions and suggestions to the JSON schema and additional suggestions for [specific styles](#). We've also contributed on to [pandoc-citeproc](#) and [pandoc](#).

After making a mistake that should have been detected by Python's `setuptools`, we [fixed](#) the [issue](#) in `setuptools` that allowed users to install improperly specified packages.

In our proposal, we included the following goals for Manubot adoption:

- At least 50 new labs using Manubot for authoring within 1 year of grant activities
- At least 200 new articles/preprints authored with Manubot within 1 year of grant activities
- At least 3 new contributors to the Manubot project within 1 year of grant activities

We use [this search](#) to return GitHub repositories that are very likely to be Manubot manuscripts. As of 2019-09-12, this returns **84 repositories**. We have reports from individuals that they use Manubot locally and/or privately, i.e. without creating a public repository. It is difficult to assess the extent of private or local usage.

While we fell short of reaching our goal of 200 manuscripts, the quality of many existing manuscripts has been high. We also have found that once a user has written one manuscript in Manubot, they are likely to continue writing subsequent manuscripts with it. Finally, interest in Manubot has spiked considerably following the publication.

Feature requests and other GitHub Issues are being opened and discussed at a healthy rate. Therefore, we feel **Manubot onboarding has progressed at a healthy pace**, as we continue to improve the experience and feature set to a point where widespread adoption makes more sense.

Going forward, we are interested in partnering with journals like the *Journal of Open Source Software* to integrate Manubot as part of their publishing workflow. These partnerships could greatly increase Manubot adoption, as could resources like the recently released Manubot catalog to drive traffic and visibility of the project.