



ILLUSTRATION BY THE PROJECT TWINS

COLLABORATIVE WRITING: BEYOND GOOGLE DOCS

A small but growing suite of tools allows researchers to author and edit scientific documents as a team, no e-mail required. **By Jeffrey M. Perkel**

Draft scientific manuscripts are typically confidential. So, when Elana Fertig was asked to take a look at an in-development paper on a functional gene-annotation strategy, she expected to receive the file in a private e-mail. What she got was a public announcement, shared on Twitter.

The paper had been written by Olga Botvinnik, a computational biologist at the Chan Zuckerberg Biohub in San Francisco, California, who is an advocate of the global movement to make research more accessible. In November 2019, as Botvinnik started preparing her paper, she decided to try this open-science

ethos out for herself. “I wanted to walk the walk of open science,” Botvinnik says.

Botvinnik managed her paper as if it were open-source software. She wrote it in a plain-text editor and placed text files alongside data sets and code for generating figures on the code-sharing site GitHub. She invited her four co-authors to submit edits using Git, software that tracks precisely how and when a file has been changed. And she used a dedicated tool called Manubot to render the document as a user-friendly manuscript, which she then published online and tweeted to the world.

Fertig, a computational biologist at the Johns Hopkins School of Medicine in

Baltimore, Maryland, says it was a “funny experience” to be tweeted an unpublished paper. “It’s a very different way of writing than the traditional academic science of not putting it out before it’s a finished product.”

Botvinnik’s manuscript was just a shell at this point: two of the figures were placeholders, and the methods section read, “We did things.” But, she says, the fact that the draft was publicly accessible made it easy to solicit feedback from co-authors and the broader community. “It’s definitely been very, very helpful to be able to show someone, ‘Here’s what I’m thinking so far. Here are some figures; here’s some text. What do you think?’”

Say ‘collaborative writing’ and most researchers probably think of Google Docs, the ubiquitous word processor that allows multiple authors to co-edit a document online in real time. But Google Docs lacks features that some scientists require, such as reference management, support for code and data and the ability to directly submit articles to journals and preprint servers.

Manubot is one of a small but growing number of tools specifically designed for collaborative writing; others include Overleaf, Authorea, Fidus Writer and Manuscripts.io. These tools not only close some of the key feature gaps, but also provide a glimpse of where scientific communication might move next.

Partners in editing

Most collaborative writing tools offer researchers a range of useful functions. Team members can keep documents private or share them with select collaborators; track changes and comment on the text; and edit documents simultaneously or asynchronously with their collaborators.

Science-focused programs supplement those with features aimed at the research community, such as built-in citation management. (Some citation managers can integrate with Google Docs using plug-ins, such as Zotero and Paperpile.) Users can generally import libraries from reference managers such as Zotero or Mendeley, or query external databases directly. The ‘cite’ button in Authorea, for example, allows users to search PubMed or CrossRef, or pull in articles by DOI or URL. In Fidus Writer, references can be added from Zotero with a simple drag-and-drop.

Manubot features cite-by-identifier, which builds bibliographies using a DOI, a PubMed or arXiv identifier or a URL, without the need for a reference manager. Inserting “@doi:10.1371/journal.pcbi.1007128” into a Manubot article, for instance, instructs the tool to find and insert a reference to the paper itself.

Botvinnik calls this approach “pretty magical”, because it circumvents the problem of researchers using (and trying to synchronize) different reference managers and libraries. “I like that I can just use the DOI and it works, and everyone else knows that there is one source of truth for the citation: the DOI,” she says.

Authorea and Overleaf support LaTeX, the typesetting language preferred by physicists, mathematicians and computer scientists. In 2017, CERN, Europe’s particle-physics laboratory near Geneva, Switzerland, adopted Overleaf as its preferred collaborative authoring platform; some 4,800 users have signed up, says CERN computing engineer Nikos Kasiousimis. LaTeX is quite an advanced system, however, so Authorea and Manubot might be better options if a simpler file format is needed. Both use the plain-text language Markdown.

Using Authorea and Manuscripts.io, authors can embed and execute software code in their

articles, and bundle figures together with the data used to create them – such features support computational reproducibility. “The intention is that you can create dynamic representations of your work, which include code, data and figures, and the narrative, all versioned together,” says Matias Piipari, founder of Manuscripts.io, which (like Authorea) is now owned by the publisher Wiley.

“As such tools gain traction, scientific articles become ever more dynamic.”

For those who prefer Google Docs, New Zealand-based Stencila is developing a plug-in that allows authors to enhance documents with executable code blocks, data tables and equations. Based on steganography, a cryptographic trick in which data are encoded in images, Stencila’s plug-in was written to “bridge that gap between the coders and the clickers”, says founder Nokome Bentley. “It’s taking the code to the environment that clickers are used to.”

Coder workflows

Manubot, by contrast, tends to appeal to coders. Developed in the laboratory of bioinformatician Casey Greene at the University of Pennsylvania in Philadelphia, the tool was designed to manage the writing of a review article on deep-learning – and coordinate its three dozen authors. The challenge: keeping track of which collaborators contributed which bit of text, line by line. “We expected to have a large number of contributors and we wanted to be able to look at the ‘atomic’ changes of one person and one group of changes,” Greene says. That is, instead of navigating a tangled mess of tracked changes, Greene wanted to be able to review each change individually, and to keep the online draft automatically up to date.

Manubot solves those problems by cobbling together various open-source tools, says Daniel Himmelstein, a Greene lab postdoc who helped to lead Manubot’s development. These include Pandoc, which provides file-conversion functionality, and GitHub Actions, which automates functions such as document creation. To set up a Manubot project, users clone a dedicated GitHub repository to their computer and modify it using a standard programming text editor, such as Emacs or SublimeText. Changes are then pushed back to GitHub, which logs them and rebuilds the document in HTML, Word or PDF format. Collaborators can modify the manuscript by submitting changes in the form of a GitHub ‘pull request’ (explore our example Manubot project at go.nature.com/39eqosg). The result is elegant, but complex.

And all of this extra functionality can require advanced programming skills. Fertig has

written grant applications using Manubot, and is comfortable with GitHub. But she won’t be using Manubot to write collaborative papers, because the level of programming it involves tends to be beyond the reach of her clinician co-authors. “There’s no way they have the bandwidth to pick up Manubot,” she says.

Easing submission

Increasingly, developers are fitting these tools with features to better encapsulate the scientific process. Some, for instance, support JATS XML, a file format commonly used in scientific publishing.

JATS XML is a structured, semantic file format that provides a rich set of metadata tags for article elements such as author names, article sections, funding sources and database accession numbers. Giuliano Maciocci, head of product and user experience at the open-access journal *eLife*, explains that the format “decouples the structure of the article from its presentation”, which makes the data easier to search, access and manipulate.

Editors typically build documents by converting author-submitted files into a format they can publish in, Maciocci says – a labour-intensive, error-prone process. To help automate this process, *eLife* is developing a tool called Libero Editor, which it hopes to release this year. Based on the Texture editor, the tool will allow *eLife* staff and authors to create and work with JATS XML documents from beginning to end. Manuscripts.io can already import JATS-formatted content, Piipari says, and it, together with Fidus Writer and the Stencila plug-in can export to that format as well.

Authorea allows authors to directly submit articles to around 41 journals and preprint archives, according to founder Alberto Pepe – and to embed interactive figures, executable code and data. Roberto Peverati, a computational chemist at the Florida Institute of Technology in Melbourne, was asked to contribute to one such journal, Wiley’s *International Journal of Quantum Chemistry*, in part to test drive Authorea. “I found it really very pleasant,” Peverati says.

As such tools gain traction, scientific articles become ever more dynamic – and responsive. On 20 March, Greene’s postdoc researcher Halie Rando created a Manubot project to try to make sense of the exploding COVID-19 literature. Within days, dozens of researchers had expressed interest in contributing. “With something as fast-moving as COVID-19, we have an urgent need for consilience, but many members of the scientific community are more isolated than usual,” Rando explains. Manubot provides a forum for these far-flung researchers to work together. “We hope to update it rapidly as new information emerges.”

Jeffrey M. Perkel is technology editor at *Nature*.