

The democratization of scientific publication

Nathan C. Sheffield^{1,2,✉}

¹Center for Public Health Genomics, University of Virginia

²Departments of Public Health Sciences, Biomedical Engineering, and Biochemistry and Molecular Genetics, University of Virginia

✉ Correspondence: nsheffield@virginia.edu

<http://databio.org/democratization-of-scientific-publication>

What's next for the revolution in scientific publishing? The scientific publishing industry is facing challenges to pillars of tradition like peer review and journal subscription. In the face of these challenges, the traditional roles of scientific publishers are deteriorating. New tools are now putting scientific publication within reach of the masses, leading to an explosion of new journals and funding models. As the material cost to publish a scientific paper approaches zero, what stops a scientist from simply self-publishing research? Here, I present my experience building a basic self-publishing system and my perspective on the changing modes of scholarly publication.

Introduction

The scientific publishing industry is grappling with existential questions about scholarly communication. What is the role of the scientific publisher in an increasingly digital world? Recent years have seen dramatic changes to the scientific publishing industry, including the rise of open access; the debut of 'mega-journals' (Björk, 2015); boycotts of traditional publishers, challenges to the impact factor (Alberts, 2013; Callaway, 2016), arguments over peer review models (Hunter, 2012); warnings about predatory publishers (Beall, 2018); and the rise and rivalry of preprint servers (Bourne *et al.*, 2017; Silva, 2017). These arguments and movements all seem to coalesce into an overarching question: what is the role of the publisher in science?

In the past, the publisher had three primary roles: filtering, formatting, and disseminating research papers. Each of these roles is beginning to erode:

Filtering. First, publishers have been the gatekeepers of scientific publication, ensuring the quality of published work. The filtering role is an organizational task to solicit reviewers and ensure research quality. This role has been challenged by increasing interest in post-publication review (Hunter, 2012) and the incredible popularity of limited-review mega-journals (Björk, 2015) and pre-print servers.

Formatting. The formatting role is perhaps the one the traditional publication model clings to most strongly: There is no doubt that a professionally formatted article is somehow more pleasant to read than a text document typeset by the author. Nevertheless, literature is increasingly consumed in reflowable web-based formats instead of rigid page-based formats, and new tools are making it easier to self-publish beautiful documents in either web or paged form.

Dissemination. Finally, traditional journals have disseminated research by selling subscriptions and printing issues, but this role has been obviously eroded by emphasis on open access and online-only distribution.

As these roles decline in importance and cost, is self-publication of beautiful research articles within reach? As more scientists realize they now have in their hands the tools to simultaneously self-publish and disseminate beautiful web pages *and* page-styled documents, could that be the proverbial nail in the coffin of the traditional publishing model?

A tale of two papers

My recent experiences with publishing papers has convinced me that we're on the verge of a *democratization of scientific publishing*. To illustrate, let me contrast my experience with two recent papers, both published in early 2018. The first was published in the *Journal of Open Source Software* (or, *JOSS*) (Sheffield *et al.*, 2018), and the second in *Bioinformatics* (Lawson *et al.*, 2018). These papers are similar; both are short *applications notes* of 1-2 pages, and both describe new R packages developed by my lab members and collaborators that are useful for bioinformatics analysis.

The submission process was also similar: in each case, we submitted the paper, which was handled by an editor who sent it out to external reviewers. In the case of *Bioinformatics*, there were 3 anonymous reviewers, and the review was handled via email; in the case of *JOSS*, there was a single, identified reviewer, and the review was handled in the open via the issue tracker on GitHub. In each case, we received feedback, addressed the concerns, improved the software and manuscript, and then resubmitted. The reviewers and editors agreed that our revisions addressed all concerns and both papers were accepted for publication.

The biggest difference between the two processes was the price of publication. For *Bioinformatics*, I elected the *open-access* option, with a price of around \$1600. For *JOSS*, all papers are open-access, but the price was \$0. Given the overall similarity in the submission process, why is there such a drastic difference in publishing price?

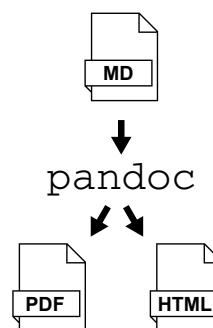
It's true that *Bioinformatics* is a traditional print journal – but the cost of printing is (theoretically) borne by the subscribers who want printed copies of the journal. In both cases reviewers are unpaid, so the imbalance in reviewer number shouldn't contribute to cost. The *Bioinformatics* manuscript was probably more thoroughly edited for grammar, which would necessitate paying a copyeditor. Another potential major cost is typesetting, which makes a polished journal article so much more appealing. However, the *JOSS* article is also quite nicely typeset.

The price difference leads to the simple question: What does it really cost a publisher to publish a scientific paper? Given the dramatic interest in starting new journals, it must be both easy and profitable to get into the publishing business. Inspired by the ability of *JOSS* to publish nice-looking articles at low cost (Smith *et al.*, 2018), I sought to see what it would take for me to set up my own mini-journal, with high-quality PDF outputs, where I could self-publish simple research results.

The journey to self-publication

Because I sought to publish my article both in reflowable text on the web as well as in a page-shaped PDF, I sought a system that could render multiple output formats from a single input source. My solution came by combining a few popular publishing tools. I had previous experience publishing web content, slides, and simple PDFs for grants using markdown, a simple human-readable text format that makes the writing portable. New tools make it possible to build nice content directly from markdown source. For example, *jekyll* produces nice web pages from markdown input; *pandoc* converts markdown into various output formats, including slideshows or basic PDFs using LaTeX (Fig. 1). Combining these things meant a single source markdown document could be rendered both on a web page as well as a PDF – the only thing I was missing was the beautifully formatted PDF template that makes you proud of your article when it comes back from the publisher. What would it take to go from my markdown-formatted basic text documents to produce *professional-looking* journal-style PDFs?

Starting with some basic TeX *pandoc* templates, I tweaked a few things: I added a two-column display, beautified the article metadata, and selected visually appealing fonts and layouts. I then built a quick shell script to run *pandoc* on my markdown-formatted blog post files, automatically rendering each post into a



A source document in markdown format can be easily converted to multiple output formats using *pandoc*.

beautified PDF article. In the end, I spent about 4 hours total putting together a template that to my eyes looks professional. With this template, I can now publish a markdown document in a beautiful journal-style PDF format with essentially no effort at all. With a few more hours of effort I was able to combine *pandoc* and *jekyll* to render a web version from the same content.

Am I now my own publisher?

Automating the process

Not only have I successfully published a document in both professional PDF and web format, but the whole thing is basically automated. It takes no extra effort to produce nice PDF output for each of my blog posts, so I've gone ahead and set up an automated system that will publish any flagged post as a PDF. Now, when I write an article that I think is worth formatting nicely into a printable PDF, I just tag that article, and suddenly users can consume it in either web or PDF format.

As an example of how this setup works, I published this very article using the system. The article is authored in markdown with standard formats for headings (e.g. # Section title), along with some minimal metadata that provides the content for the title, author, date, class, and abstract sections in a standard markdown yaml metadata header block. Now with my *.tex* template, which I've made [publicly available](#), and relying on a few other tools like *pandoc-citeproc* and *BibTeX*, a simple *pandoc* command renders a beautifully formatted PDF:

```
pandoc ${post} \  
  -o pdfs/${filename}.pdf \  
  --filter pandoc-crossref \  
  --filter pandoc-citeproc \  
  --template $textemplate \  
  --bibliography $bib \  
  --csl $csl
```

The writing on the wall

Altogether, scientific publishing is a multi-billion dollar industry, with a recent study pinning the price of publishing typical paper at around \$3,000 for an open access, online-only publication, with some journals charging even more (Noorden, 2013). Are these numbers warranted? The traditional roles of the publisher are all being challenged by new technologies and philosophies. The actual incurred *cost* of publishing has decreased to the point that it's now feasible for an individual, in a matter of hours and at no monetary expense, to effectively set up an independent journal.

In a world where real costs of publishing approach zero, how can the price of publication remain so high? It is due to the *reputation capital* built by publication houses over years of scientific publishing. This capital is manifest as readership and impact factor, which is entrenched as a currency in academia. But while traditions like this die hard, it's also true that the publishing times are changing (Chawla, 2015); as the pillars of publishing erode, publishers must reinvent themselves. Perhaps these changes will lead to decreased publishing prices at the entrenched publishing houses. Or perhaps publishers will create new value as they redefine the role of a publisher in the 21st century.

Or... maybe researchers will bypass publishers entirely with self-publication, and scientific publication will become truly democratic.

References

Alberts,B. (2013) Impact factor distortions. *Science*, **340**, 787–787.

Beall,J. (2018) Scientific soundness and the problem of predatory journals. *Pseudoscience: The Conspiracy Against Science*, 283.

Björk,B.-C. (2015) Have the “mega-journals” reached the limits to growth? *PeerJ*, **3**, e981.

Bourne,P.E. *et al.* (2017) Ten simple rules to consider regarding preprint submission. *PLOS Computational Biology*, **13**, e1005473.

Callaway,E. (2016) Beat it, impact factor! Publishing elite turns against controversial metric. *Nature*, **535**, 210–211.

Chawla,D.S. (2015) What's in a journal name? *Nature*, **528**, 311–311.

Hunter,J. (2012) Post-publication peer review: Opening up scientific conversation. *Frontiers in Computational Neuroscience*, **6**.

Lawson,J. *et al.* (2018) MIRA: An R package for DNA methylation-based inference of regulatory activity. *Bioinformatics*, **bty083**.

Noorden,R.V. (2013) Open access: The true cost of science publishing. *Nature*, **495**, 426–429.

Sheffield,N.C. *et al.* (2018) simpleCache: R caching for reproducible, distributed, large-scale projects. *The Journal of Open Source Software*, **3**, 463.

Silva,J.A.T. da (2017) The preprint wars. *AME Medical Journal*, **2**, 74–74.

Smith,A.M. *et al.* (2018) Journal of open source software (JOSS): Design and first-year review. *PeerJ Computer Science*, **4**, e147.