

Science Publishing's Napster Moment and the Coming Youtubes of Science

Josh Nicholson¹ and Alberto Pepe¹

¹Authorea Team

April 28, 2020

Research is really f**king important.

This statement is almost self-evident by the fact that you're reading this online. From research has come the web, life-saving vaccines, pasteurization, and countless other advancements. In other words, you can look at cat gifs all day because of research, you're alive because of research, and you can safely add milk to your coffee or tea without contracting some disease, because of research. But how research is done today is being stymied by how it is being communicated. Most research is locked behind expensive paywalls (Björk et al., 2010), is not communicated to the public or scientific community until months or years after the experiments are done (Himmelstein and Powell, 2016), is biased in how it is reported - only "positive" results are typically published (Ahmed et al., 2012), does not supply the underlying data to major studies (Alsheikh-Ali et al., 2011), and has been found to be irreproducible at alarming rates (Begley and Ellis, 2012).

Why is science communication so broken?

Many would blame the fault of old profit-hungry publishers, like Elsevier, and in many respects, that blame is deserved. However, here's a different hypothesis: what is holding us back from a real shift in the research communication industry is not Elsevier, it's Microsoft Word. Yes, Word, the same application that introduced us to Clippy is the real impediment to effective communication in research.

Today, researchers are judged by their publications, both in terms of quantity and prestige. Accordingly, researchers write up their documents and send them to the most prestigious journals they think they can publish in. The journals, owned by large multinational corporations, charge researchers to publish their work and then again charge institutions to subscribe to the content. Such subscriptions can run into the many millions of dollars per year per institution (Lawson and Meghreblian, 2015) with individual access costing \$30-60 per article.

The system and process for publishing and disseminating research is inimical to scientific advancement and accordingly Open Access and Open Science movements have made big steps towards improving how research is disseminated. Recently, Germany, Peru, and Taiwan have boycotted subscriptions to Elsevier (Schiermeier and Mega, 2016) and an ongoing boycott to publish or review for certain publishers has accumulated the signatures of 16,493 researchers and counting. New developments such as Sci-hub, have helped to make research accessible, albeit illegally. While regarded as a victory by many, the Sci-hub approach is not *the* solution that researchers are hoping for as it is built on an illegal system of exchanging copyrighted content and bypassing publisher paywalls (Priego, 2016).

The most interesting technologist view of the matter is that the real culprit for keeping science closed isn't actually the oligopoly of publishers (Larivière et al., 2015)– after all, they're for-profit companies trying to run businesses and they're entitled to do any legal thing that helps them deliver value to shareholders. We suggest that a concrete solution for true open access is already out there and it's 100% legal.

What is the best solution to truly and legally open access to research?

The solution is publishing preprints – the last version of a paper that belongs to an author before it is submitted to a journal for peer review. Unlike other industries (e.g. literature, music, film, etc.), in research, the preprint version copyright is legally held by the author, even after publication of the work in a journal. Preprints are rapidly gaining adoption in the scientific community, with a couple of preprint servers (e.g. arXiv which is run by Cornell University and is primarily for physics papers, and bioRxiv which is similarly for biology papers) receiving thousands of preprints per month. Some of the multinationals are responding with threats against authors not to publish (or post) preprints. However they are being met with fierce opposition from the scientific community, and the tide seems to be turning. Multinationals are now under immense pressure not just from authors in the scientific community, but increasingly from the sources of public and private funding for the actual research. Some organizations are even mandating preprints as a condition of funding. But what is holding back preprints and in general a better way for Authors to have more control of their research? We think the inability for scientists to independently produce and disseminate their work is a major impediment and at the heart of that of that problem is how scientists write.

How can Microsoft Word harm scientific communication?

Whereas other industries, like the music industry, have been radically transformed and accelerated by providing creators with powerful tools like Youtube, there is no parallel in research. Researchers are reliant upon publishers to get their ideas out and because of this, they are forced into an antiquated system that has remained largely stagnant since it's inception over 350 years ago. Whereas a minority of researchers in math-heavy disciplines write using typesetting formats like LaTeX, the large majority of researchers (~82%) write their documents in Microsoft Word (Brischoux and Legagneux, 2009). Word is easy to use for basic editing but is essentially incompatible with online publishing. Word was created for the personal computer: offline, single-author use. Also, it was not built with scientific research in mind - as such, it lacks support for complex objects like tables and math, data, and code. All in all, Word is extraordinarily feature-poor compared to what we can accomplish today with an online collaborative platform. Because publishers have traditionally accepted manuscripts formatted in Word, and because they consistently fail to truly innovate from a technological standpoint, millions of researchers find themselves using Word. In turn, the research they publish is non-discoverable on the web, data-less, non-actionable, not reusable and, most likely, behind a paywall.

What does the scientific communication ecosystem of the future look like?

What is needed is a web-first solution. Research articles should be available on distinct web pages, Wikipedia style. Real data should live underneath the tables and figures. Research needs to finally be machine readable (instead of just tagged with keywords) so that it may be found and processed by search engines and machines. Modern research also deserves to have rich media enhancement – visualizations, videos, and other forms of rich data in the document itself. All told, researchers need to be able to disseminate their ideas in a web first world, while playing the “Journal game” as long as it exists. Our particular dream (www.authorea.com) is to construct a democratic platform for scientific research – a vast organizational space for scientists to read

and contribute cutting edge science. There is a new class of startups out there doing similar things with the research cycle, and we feel like there is a real and urgent demand for solutions right now in research.

References

- I. Ahmed, A. J. Sutton, and R. D. Riley. Assessment of publication bias selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *BMJ*, 344(jan03 1):d7762–d7762, jan 2012. doi: 10.1136/bmj.d7762. URL <http://dx.doi.org/10.1136/bmj.d7762>.
- Alawi A. Alsheikh-Ali, Waqas Qureshi, Mouaz H. Al-Mallah, and John P. A. Ioannidis. Public Availability of Published Research Data in High-Impact Journals. *PLoS ONE*, 6(9):e24357, sep 2011. doi: 10.1371/journal.pone.0024357. URL <http://dx.doi.org/10.1371/journal.pone.0024357>.
- C. Glenn Begley and Lee M. Ellis. Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391):531–533, mar 2012. doi: 10.1038/483531a. URL <http://dx.doi.org/10.1038/483531a>.
- Bo-Christer Björk, Patrik Welling, Mikael Laakso, Peter Majlender, Turid Hedlund, and Guni Gunason. Open Access to the Scientific Journal Literature: Situation 2009. *PLoS ONE*, 5(6):e11273, jun 2010. doi: 10.1371/journal.pone.0011273. URL <http://dx.doi.org/10.1371/journal.pone.0011273>.
- François Brischoux and Pierre Legagneux. Don’t Format Manuscripts. *The Scientist*, 23(7):24, 2009.
- Daniel S. Himmelstein and Kendall Powell. Analysis for the history of publishing delays blog post v1.0, 2016. URL <https://doi.org/10.5281/zenodo.45516>.
- Vincent Larivière, Stefanie Haustein, and Philippe Mongeon. The Oligopoly of Academic Publishers in the Digital Era. *PLoS ONE*, 10(6):e0127502, jun 2015. doi: 10.1371/journal.pone.0127502. URL <http://dx.doi.org/10.1371/journal.pone.0127502>.
- Stuart Lawson and Ben Meghreblian. Journal subscription expenditure of UK higher education institutions. *F1000Research*, jul 2015. doi: 10.12688/f1000research.5706.3. URL <http://dx.doi.org/10.12688/f1000research.5706.3>.
- Ernesto Priego. Signal, Not Solution: Notes on Why Sci-Hub Is Not Opening Access, 2016. URL <http://dx.doi.org/10.15200/winn.145624.49417>.
- Quirin Schiermeier and Emiliano Rodríguez Mega. Scientists in Germany Peru and Taiwan to lose access to Elsevier journals. *Nature*, dec 2016. doi: 10.1038/nature.2016.21223. URL <http://dx.doi.org/10.1038/nature.2016.21223>.