

2016

Everything Evolves, Even Publishing

Jason Hoyt
PeerJ, jason@peerj.com

Peter Binfield
PeerJ, pete@peerj.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Hoyt, Jason and Binfield, Peter (2018) "Everything Evolves, Even Publishing," *Against the Grain*: Vol. 28: Iss. 1, Article 12.
DOI: <https://doi.org/10.7771/2380-176X.7270>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

How and Why Data Repositories ... from page 24

As part of **Figshare's** partnership with **Nature Publishing Group** and their journal *Scientific Data*, we've been able to analyze user behaviour and preferences. *Scientific Data* ask researchers to place data in structured data repositories, institutional repositories or both when suitable ones exist. Tellingly, over 30% of data submissions were made to **Figshare**, making it the most used repository. We know from this that the majority of researchers require an unstructured repository for their data. The extent to which this will change over time as codification and structuring efforts proceed is arguable. It is our opinion that there will always be a strong need for unstructured repositories because it is the nature of research that many experiments and techniques are novel and unique.

Where Does this Leave Us?

It has taken longer than expected for the promise of the digital age to begin to make a real difference to the way scholars communicate their work. The persistence of traditional measures of quality are the most likely explanation for academia's apparent conservatism, but with funding bodies increasingly encouraging and mandating the sharing of data, we are finally seeing diversification of what is considered legitimate scholarship.

The publishing industry has made strides over the last decade or so to integrate with institutional, funder and community based repositories. Together with groups interested in the standardization of data formats, a lot of progress has been made to codify formats in many fields. There remains, however a large quantity of data on researchers' hard drives and servers that don't fit into easily standardized formats because the techniques are either new or unique.

There are still many open questions in data publishing, from how to deal with embargoes or sensitive data to how best to assess the quality of the diverse range of digital research outputs. The field of data publishing is still in its formative stages and represents an opportunity for both publishers and libraries to help academics adapt to new requirements. 🐼

Endnotes

1. **Piwowar H. A., Day R. S., Fridsma D. B.** (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. doi:10.1371/journal.pone.0000308
2. **Meadows A.** (2011) *Scholarly Kitchen*. <http://scholarlykitchen.sspnet.org/2014/11/11/to-share-or-not-to-share-that-is-the-research-data-question/>
3. **Schaffer, T. and Jackson, K. M.** 2004. The use of online supplementary material in high-impact scientific journals. *Science & Technology Libraries* 25(1/2):73-85.
4. The **Royal Society Science Policy Centre** report: *Science as an open enterprise* (2012) https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.
5. **Valen D. and Blanchat K.** (2015) **Figshare** Blog https://figshare.com/articles/Overview_of_OSTP_Responses/1367165.
6. **Stemwedel J. D.** (2015) *Forbes* <http://www.forbes.com/sites/janestemwedel/2015/06/01/reasons-to-keep-discussing-the-lacour-and-green-retraction/>.
7. **Laure Haak L.** (2015) **ORCID** blog <https://orcid.org/blog/2015/10/26/auto-update-has-arrived-orcid-records-move-next-level>.
8. **Jones P. B., Wesolek A., Scherer D., Watkinson A.** (2015). A Game of Spot the Difference: Librarians, Repository Managers and Publishers. Presentation slides. http://works.bepress.com/andrew_wesolek/30/
9. **Giesecke J.** (2011), "Institutional Repositories: Keys to Success" Faculty Publications, **UNL** Libraries. Paper 255. <http://digitalcommons.unl.edu/libraryscience/255>

Everything Evolves, Even Publishing

by **Jason Hoyt, PhD.** (CEO and Co-founder, PeerJ) <jason@peerj.com>

and **Peter Binfield, PhD.** (Publisher and co-founder, PeerJ) <pete@peerj.com>

We sometimes hear that for all the promise of the Internet, it is a shame that it has yet to impact scholarly communication in the same way it has other industries. One could argue this point quite effectively: prestige still dominates; the journal name matters just as much as it always has; the same legacy publishers still control most of the literature; Open Access is just a small fraction of all articles, etc., etc. Meanwhile, in other industries it is easy to spot how the old guards have changed and new names have sprung up: **Google, Wikipedia, Amazon, Uber** and **Facebook** to name just a few.

On the other hand, does anyone believe Open Access is going away? Will data not

become more widely available? Will tools to make publishing faster never be developed? Why have "megajournals" appeared in the past ten years and not just survived, but become the future revenue model for new and old publishers? Why are scholarly societies struggling after decades/centuries of thriving? Why are governments and funders making Open Access mandates? These events contradict the notion that the Internet hasn't changed things in an "unmovable" 300 year-old industry. Indeed, the evidence actually suggests that we are in the midst of a change so expansive that we don't quite know how to adapt to it.

We take comfort in the way things worked in the past, as they had slowly developed in manageable timetables over the 20th century. There was certainty in how to communicate science, who to trust, or what to do for academic career progression. We now live in an era with an alluring future, but one that raises new concerns:

How will we fund scholarly output? How much

should we make open, and how? Is publishing Open Access a bet on the future, or will it negatively affect my students or my career?

What the last ten years or so have done is to open our minds to questions that many of us never anticipated having to find solutions for. It could be argued that just as the Internet has made us more globally aware, so academia has grown more concerned with its impacts outside of the ivory tower. The decentralization that occurred with the World Wide Web makes it clear how we affect those around us, and this has influenced our professional lives in a similar way. It's not that scientists are only just now waking up to the fact that they can be open, they just didn't realize it was possible until recently. Our policies and infrastructures are unprepared for these changes, just as much as our readiness to leave the comfort of the past.

There Would be no Open or Mega-Journals without the Internet

Just as the printed journal was a forgone conclusion of the printing press, so too was Open Access and the megajournal a natural by-product of the Internet. Perhaps someone

continued on page 26



before the Internet's arrival envisioned a world of Open Access, but it is more likely that no one had conceived what the potential for scholarly communication would be even as recently as 1990. The technology of the time didn't allow for anything other than the printed article with not just limits on article length, but also to what type of research could be done.

The same advancements that brought forth the Internet, for example "Moore's Law," also brought us more powerful computational resources and tools. This led us to new ideas and new science, which in turn made big data science a "thing" and meant that what was previously considered adequate, the printed article, was no longer a sufficiently-sized container in the Internet era.

The Internet also made us rethink who research should be serving. With printed literature, the boundaries of information access seemed clear — distributing a printed article to everyone in the world just wasn't thinkable. But now we have entered a world in which anyone with access to a computer and the Internet could conceivably retrieve information instantly and cheaply. Unlike a printed article, duplicating information stored as bits was virtually free.

Indeed, the **Budapest Open Access Initiative** and the definition of Open Access arose out of this reflection of what the Internet meant.

In summary, the Internet changed science and our expectations of scholarly communication in three fundamental ways:

1. Distribution has become commoditized. Articles of any length, and their corresponding journals, can be distributed for the same cost as smaller articles.
2. The same technology that made the Internet possible also started to generate new types of research, output formats, and large amounts of data.
3. "Free access" to research for anyone was a possibility.

The first two changes have given rise to Megajournals, whilst the third represents Open Access. We put quotes around "Free access" because it actually refers to two key points of Open Access. First, that there is no financial barrier to obtaining the research article (what is sometimes referred to as "free, as in 'free beer'"). And secondly, that there are no legal or technical restrictions to reading, downloading, or reusing the research conclusions. For example, in the case of CC BY distribution, copyright remains with the author, but anyone is allowed to download and reuse the article.

These two key points of the Open Access definition present a problem, however. That is, how do we find a sustainable solution to these lofty ideals?

Toward Sustaining Open Access

While the Internet has reduced the cost to make duplicate copies of a research article and instantly deliver it to the other side of the

planet, there are still costs upstream. Some of this is still an expectation, in many disciplines, that the finished product look like we've always seen in the printed format — nice typesetting, well-designed, etc. High quality production and typesetting still costs money.

Other costs are the long-term considerations for archiving. In the event that a journal should disappear, then there needs to be plans in place to preserve the content indefinitely, and so third-parties (such as **Portico** or **CLOCKSS**) are used and paid to ensure published research doesn't disappear along with its journal.

Then there is the human labor cost. While reviewing is usually done on a volunteer basis and organized by an Academic Editor, who is also usually a volunteer, the system behind that is complex. Certainly, a handful of academics could and do get together to produce some journals without any paid employees, but this is very rare. Ensuring a smooth, speedy, and standards-compliant process at scale still requires paying a staff. Authors need to be checked; reviewers need to be chased; editorial queries need to be resolved.

All of these factors add up to a non-trivial amount. And even in venues such as arXiv that have no expectation of typesetting, proofing, long-term archiving, and no peer-review, there are large costs reaching nearly \$1M annually.

These costs have meant that to reach the goal of reading and downloading for free, there had to be money coming from some other source. While Open Access says nothing about the financial model, it has become common to associate most peer-reviewed Open Access articles with the "Gold OA" model. Popularized by **BioMedCentral**, in the Gold OA model the publication charges are paid for by the author in some way (either personally, via a grant, or through their institution).

A "hybrid" model has also appeared in traditional subscription journals — where an article in a pay-walled journal can be made Open Access by paying the article charge, however, other articles still remain behind the pay-wall. This model has been met with some controversy, as there are concerns that publishers are "double dipping" by taking both subscription money and the Open Access article fee.

At **PeerJ** we've developed another path, which doesn't depend on a per article charge, but rather is a membership per author (though **PeerJ** also has the traditional per article pricing as well). The membership model is a refinement that helps to further reduce the financial burden toward sustainable Open Access. It isn't the only thing contributing to lower OA costs, technical innovation plays a large part, but it does show that publishing high quality Open Access can feasibly drop to a very low cost.

Going hand-in-hand with Open Access and the Internet was the realization that *what the journal* can be changes when there are no space constraints. This is the "megajournal."

The Megajournal and Publisher Evolution

As mentioned, the cost to reproduce and distribute digital bits in the Internet era is trivial; it therefore makes sense that the cost of

displaying a longer article is also trivial (other than upstream and archiving costs previously discussed). It also follows that if you have a business model that can pay the cost of each individual article (rather than pay at the journal level), then a "journal" need not be limited to a set number of articles per issue. Thus, it was only a matter of time before a journal arrived without such constraints.

This journal was *PLOS ONE*, of course. In its first year it published more than 1,200 articles. Within six years it was publishing more than 30,000 annually — as a single journal.

Part of the success to this was not just the format change, but an editorial policy of not evaluating for novelty or importance (and instead focussing on sound science). Because articles are not rejected on "spurious grounds," the acceptance rate increases, and thus publication numbers increase as well, giving rise to "the megajournal."

The megajournal *PLOS ONE* turned out to not just be successful in publishing a large part of the STM literature (nearly 3% of it annually), but it was also a financial success. It more than subsidized the other Open Access journals in the **PLOS** organization's portfolio that were running a traditional limited article issue — despite appearing online only. For the first time, a path towards a sustainable Open Access future started to appear.

The megajournal model has been so financially successful that nearly every major publisher has now started an Open Access megajournal (including **PeerJ**, of course). And so, while traditional publishers still run subscription-based journals, the Open Access model is rapidly becoming their fastest growing market.

With every publisher now entering the Open Access megajournal game, a new type of competition has entered the academic scholarly market. Prestige still dominates, however, megajournals now also need to appeal to the individual author who decides where to publish and pay. The "author experience" matters now, more than ever before.

The core author experience involves the submission platform for any journal. Under the subscription model, where prestige dominates, authors are more willing to put up with difficult submission workflows, and software (along with unpleasant or slow peer-review). It made sense that this non-core facility of subscription journals would be outsourced. That has changed with the megajournal competition and Open Access.

PeerJ was the first to recognize how "core" the submission experience is to attracting authors in the megajournal world, which is why it built the entire workflow in-house rather than licensing an outside vendor product. Other publishers have now been following suit, and naturally that is also flowing into the subscription submission systems as well.

In just the last few years, the core competencies needed by a modern academic publisher have drastically changed. It now makes sense to have in-house expertise in technology and user experience. The megajournal landscape

continued on page 28

is re-shaping user expectations, much like what the iPhone and Google's "material design" have done.

With the rise of the megajournal and Open Access, however, we're now recognizing a new issue—journal prestige is a holdover from the past...

The Conundrum of the Megajournal, Open Access, and Prestige

It is through the historical artefact of print that we developed the still current mechanisms of funding, tenure, and other facets of the academic world. In the resource-limited era of print, it made *some* sense to use the journal as a proxy for quality of the individual article. This was further exacerbated by a growing reliance on the Impact Factor in the late 1900s.

Individuals and organizations could afford to purchase, deliver, and find only a limited number of articles in the print and pre-Internet world. Journal names, and the "filter" they represented carried a lot of weight. Those limitations do not exist now as search engines, recommender systems, and boundless access to Open Access literature means we can virtually filter every journal. The only limitation is whether the article is Open Access and the quality of the filtering process.

Attitudes are shifting though — the brand name journals are no longer always the first choice for scientists, as Open Access is now frequently more important. Funders, and even entire countries, are also mandating

the research they fund find a home in Open Access venues.

And the traditional brand-name journals are increasingly failing due to the increasing pressure to always publish what is perceived as the most novel findings. These policies result in more retractions in the "top" journals.

Statistically, it makes sense that the best research and best authors are more and more likely to be found in megajournals and Open Access venues as they account for more than 10% of the literature.

However, hiring, tenure, and grant committees are struggling with these changes. For years they have relied upon just the journal name and, by extension, the Impact Factor to make decisions. The problem isn't so much that good research can't be spotted in Open Access journals, but it seems to be the uncomfortable acknowledgment, due to tradition, that good research isn't just published in "brand name" high impact journals anymore.

This isn't a problem necessarily solved by technology either. Even with the best of altmetrics, existing or yet to be innovated, we will still have this perception problem with Open Access and megajournals. These types of problems require a different set of solutions: research, policy, and education.

Open Access and megajournals have become a valuable asset and look like the future of scholarly communication. However, we recognize that comfortable traditions are being upended with these changes, and so we propose three strategies to smooth this transition:

1. Top-level research is needed to understand these changes more

thoroughly. For example, how are committees handling these changes — what examples of successful transitions are there, how were they implemented, or what else can be learned from them? What are the impacts of making decisions still based on the print era information? And how are organizations and individuals transitioning to fund Open Access?

2. From that research we should be able to start developing new policies at different governance levels to aid in the transition. We need to ensure better decisions are being made at the author and article levels, and that Open Access continues to have a sustainable future.
3. Finally, educational and influencer campaigns are a must if we are going to upend perceptions of where the best research is located and how Open Access can and should affect career progression and decisions. Senior researchers are a powerful influence, and should be encouraged to send their best work to Open Access journals.

In conclusion, the Internet has had a profound impact on scholarly publishing. It causes us to question much of how we decide what to fund, hire, read, and where to publish. There are many unanswered questions that will require a concerted strategy to understand and implement solutions in the Open Access world that we now live in. 🐙

Adaptations in Publishing — Publishers and Librarians Advancing Research

by **Maggie Farrell** (Dean of Libraries, Clemson University) <maggie4@clemson.edu>

and **Alicia Wise** (Director of Access and Policy, Elsevier) <a.wise@elsevier.com>

Introduction

Research is improved when researchers are able to connect with other individuals to share results, concepts, theories, data, and disagreements. Researchers thrive in an open environment where theories and results are readily available. Researchers are motivated by scientific discovery, access for their informational needs, and promotion of their own and other's works. Non-researchers also gain expertise and knowledge when they have access to current research and science.

With the advancement of technology, publishers and librarians have an opportunity to create new environments that facilitate sharing and communication during the research process as well as access to final research results, supporting data, reviews, and ongoing work based on the research results. Researchers

themselves are also involved in creating new environments, for example by engaging social media to connect in new ways with new audiences. These new environments enable publishers, librarians, and researchers to work together in new ways to advance research and make it more efficient and impactful. Technology also provides an opportunity for publishers and librarians to create a new relationship in support of mutual goals of expanding research and supporting researchers throughout the research process.

Overview of Ways in which Publishers Support the Research Process

Publishers provide access to, and quality assurance of, content, data, systems, tools and analytics that help universities and researchers define, manage and achieve their desired

research outcomes, and to promote awareness of their impact. Well-designed technology underpins all of these services — and the largest publishers utilize high performance computing clusters to support robust analytics and big data processing. And publishing at its heart is a very social endeavor, relying as it does on human relationships. For example, here at **Elsevier** we work closely with over 7,000 editors, 70,000 editorial board members, 300,000 reviewers, and more than 1,000,000 corresponding authors every year. By working with librarians, more than 12,000,000 researchers at thousands of institutions and companies worldwide have access to our products.

Authors too want to be able to disseminate and share their research, and publishers actively provide tools and services to enable

continued on page 29