There is a pathetic lack of functionality in scholarly publishing. We must end for-profit publishing and allow libraries to make available the works of their scholars for all

by Blog Admin   November 9, 2011

Publicly-funded science is suffering but academia must embrace technology before it can deliver its full potential to scientists, policy-makers and the public. **Björn Brembs** *argues that the sum made by for-profit publishers would be more than enough to establish a freely accessible infrastructure that would ensure scholarly knowledge and research remain in the hands of libraries, and the public.*

Can we save scholarly publishing? Yes we can, but it sure takes an optimist to believe it. Let's just take, as a case study, one of the many tasks scholarly publishing fails miserably at: allowing the scientists to stay on top of the scientific discoveries in their particular field. Scholarly publishing used to be about scientists communicating their discoveries to other scientists. Today, these discoveries are buried somewhere among 24,000 journals – most of which cannot be accessed by the individual scientist because his or her institution does not subscribe to them. The journal hierarchy that has established itself over the last four to five decades is also useless: sometimes, the 'high-ranking' journals publish some of the most discredited and easily refutable papers and some of the most obscure journals feature some extremely relevant information. Thus, any rule of only following some journals results in a waste of time at best and missing relevant information at worst.

Some of my colleagues have asked me if I don't see a reflection of the journal hierarchy in the papers they publish, and if I wouldn't review differently for 'higher' journals. For both questions, I have to answer with a resounding 'no', and the data backs me up: there is very little obvious correlation between an article and the rank of the journal it was published in, but a rather strong correlation between the number of retractions in a journal and its rank. And as if these correlations weren't enough to convince my colleagues that any impression of paper-rank inferred by container-rank are not based in any evidence, the dominant metric by which this journal rank is established, Thomson Reuters' 'Impact Factor' (IF) is so embarrassingly flawed, it boggles the mind that any scientist can utter these two words without blushing: the IF is negotiable and doesn't reflect actual citation counts; the IF cannot be reproduced, even if it reflected actual citations; and the IF is not statistically sound, even if it were reproducible and reflected actual citations.

There is thus more than ample evidence in favor of the hypothesis that *where* something is published is actually quite irrelevant, and no evidence that I know of contradicting it.

Staying relevant amongst 24,000 journals

If it is indeed irrelevant where something is published, doesn't that mean we have to somehow screen the 24,000 journals with their 2 million papers every year for the comparatively few papers that actually are relevant to the research of the working scientist? Indeed, that is the case. However, these journals are not on Google. A few thousand of them are on Thomson Reuters' "Web of Science", a few thousand on "PubMed", a few thousand on "Google Scholar", a few thousand on "Microsoft Academic Search" and a few thousand here and there on some other specialized search engines that I wouldn't know as they're outside of biomedicine. The degree of overlap between these information silos varies widely. Consequentially, it differs quite a lot what people do to stay current. Here's what I have evolved to do (and the process keeps changing):

1. Pick ten, fifteen or so journals I have access to and which have a reasonable chance of publishing something important in my field and read their tables of contents religiously

2. Scan for any citation alerts: if they cite me, it must be relevant!

3. Go through stored PubMed keyword searches (even though PubMed is 4 weeks behind publication date)

4. Read F1000 alerts

5. Subscribe to 3-4 mailing lists on which some helpful person posts press-releases in close enough fields

6. Screen citeUlike recommendations and FriendFeed subscriptions or Tweets for interesting papers

7. Subscribe to science news wires

8. Listen to science podcasts

I'd guesstimate that this takes about 12-14h per week just to find the papers. Usually, that amount of time spent searching leaves me with no time to actually read the things I found!
And the system, as complicated as it is, isn't even doing a good job: just the other day I was alerted to an extremely important paper for my research, in a high-ranking journal, by a colleague by pure accident: the title didn't look relevant, the authors were not triggering the keywords I had saved (because they used a different terminology) and were not citing any of our papers, because they worked in related field and probably didn't know that we were doing related work either. How many other relevant papers have I missed in that way? And I know I scan hundreds of irrelevant paper-titles each day.

Some people don't even try anymore. Our professor emeritus at the institute once admitted: *"I don't really follow the literature anymore. If there's something really important, it'll find its way to me."*

This is only one of many other tasks. Scholarly publishing is also pretty bad at connecting the actual data with the text describing the experiments and their interpretation. Scholarly publishing often can't even distinguish between James Smith and John Smith and easily thinks a married scientist who changed their name is a different person. If you click on a phrase that says "experiments were conducted as previously described", very often, nothing happens. If you're lucky, you see the item in the reference list. If you're very lucky, that item will contain a link to the paper they cited. If you're obscenely lucky, that link points to a service and maybe even the paper in a journal your institution subscribes to. In no case ever, any of these links will get you precisely to the section in the cited paper that describes the experiment. The first famous demonstration of hyperlink technology is from 1968. More than four decades later, scholarly publishing has still to embrace that technology. Scholarly publishing relies on hashtags such as #icanhazpdf on Twitter to get scientists access to papers (in PDF format, no hyperlinks!).

For-profit publishing versus open access

You would really be forgiven if you were to start crying at that enumeration of pathetic lack of functionality. But it gets even worse: the multi-national corporations that control scholarly publishing actually siphon off billions of dollars from this neanderthal enterprise, at profit margins exceeding 30%. In other words, not only do publicly funded scientists and science suffer, the taxpayer is even lining the pockets of the international shareholders who are holding them hostage: "give me your money or not even your doctors will get access to the information that could save your lives – let alone you!"

One of the largest publishers in the business, Elsevier, notorious for once publishing a set of fake journals in the disguise of peer-reviewed literature, with the intent of marketing pharmaceuticals to doctors, is currently making more than 800 million Euros in annual profits. This profit of one single for-profit publisher would be enough to buy 60% of all the papers published every year and make them accessible for everyone. Combined with just the profits from scholarly publishing of one of the other big players, let's say Thomson Reuters (mainly from its "Web of Knowledge"), there would be enough money to make every single publication open access, every single year from now on (and wouldn't even touch the profits of the other publishers).

Time to invest in a new model

And this brings me to the point why scholarly publishing can be saved: depending on what sources you use and which profits are counted, the for-profit scholarly publishing sector rakes in an annual profit of anywhere between 2 and 4 billion Euros in largely taxpayer funds. This is more than enough money not only to make all the publicly funded research accessible to the taxpayer that funded it, but there would be plenty left to invest in infrastructure to develop a smart alerting service where I would spend one hour a week searching for the literature and ten hours reading it. There would be money left over to invest in archiving strategies to make scholarly knowledge last beyond financial catastrophes. There would be a completely new sense of purpose bestowed on the one institution that has hundreds and hundreds of years of experience in archiving scholarly output and making it accessible: the university library.

Yes, I suggest to get rid of for-profit scholarly publishing altogether and let the libraries again host the work of their scholars, as it once was. This new, decentralized, federated database of scholarly work would be all the below and more:

- A single semantic, decentralized, federated database of literature and data
- Personalized filtering, sorting and discovery
- Peer-review administrated by an independent body
- Link typology for text/text, data/data and text/data links ("citations")
- Semantic Text/Datamining
- All the metrics you (don't) want (but need)
- Tagging, bookmarking, etc.
- Unique contributor IDs with attribution/reputation system (teaching, reviewing, curating, blogging, etc.)
- Technically feasible today

Conclusion

Scholarly publishing is badly broken, but not beyond repair. The exorbitant profits that corporate publishers currently extract from the taxpayer provide an enticing avenue out of the current misery. If university libraries were to cancel or reduce subscription contracts with corporate publishers in a step-wise fashion and, importantly, in excess of what budget constraints already force them to do, they would have increasingly larger funds at their disposal. These funds would, at the end of that probably many year long process, all else remaining equal, amount to approx. 2-4 billion dollars per annum. These funds could, from the very first year on, be used to invest in the necessary infrastructure which would provide much of the functionality which scholarly publishing is so bitterly lacking today. I predict that the ensuing lack of access will win support rather than opposition from the affected faculty, if some of the funds are diverted towards intermediary open access funding or color/page charges.

Related posts:

1. [By championing open access publishing, the academic community can bring us closer to making research available to all.](#)
2. [Continual publishing across journals, blogs and social media maximises impact by increasing the size of the 'academic footprint'.](#)
3. [Universities are increasingly moving towards recognising digital scholarship despite conflicting messages that favour traditional publishing in journals](#)
4. [Open access repositories are beginning to push academic publishers off their previously unreachable perch.](#)
5. [Social media is inherently a system of peer evaluation and is changing the way scholars disseminate their research, raising questions about the way we evaluate academic authority](#)