

PLoS One: Interview with Peter Binfield

Martin Fenner, Gobbledygook

August 15, 2009

At SciBar Camp Palo Alto last month, Peter Binfield from PLoS ONE gave a very interesting presentation on Article-level metrics from the PLoS perspective. Particularly interesting was his announcement that PLoS journals will provide usage data (HTML pageviews, PDF and XML downloads) for all their articles in September. Usage data, like all measures of scientific impact, have their problems, but they are a welcome addition to citation-based metrics.

I've interviewed Pete to ask him not only about article-level metrics, but also about the publishing model of PLoS ONE and how these two relate to each other.

1. Can you describe what PLoS ONE is and does?

PLoS ONE is an Open Access, scholarly, peer reviewed journal for all of science. We will be three years old in December 2009, but already we are the third largest journal (of any type) in the world, publishing approximately 4,600 articles in 2009 alone and almost doubling in volume every year. We are online only (publishing in HTML, XML and PDF) and we publish daily (about 20 articles a day at present). In my view PLoS ONE is the most dynamic, innovative and exciting journal in the world, and I am proud to work on it.

In many ways PLoS ONE operates like any other journal however it diverges in several important respects. The founding principle of PLoS ONE was that there are certain aspects of publishing which are best conducted pre-publication and certain aspects which are best conducted post-publication. The advent of online publishing has allowed us to take a step back and re-evaluate these aspects of how we publish research, without the burden of centuries of tradition. In this way, we have been able to experiment with new ways of doing things which may result in dramatic improvements in the entire process of scholarly publication.

The most important thing which has come out of this premise is that unlike almost every other journal in the world, we make no judgment call whatsoever on the “impact” or “significance” or “interest level” of any submission. What this means is that if an article appropriately reports on well-conducted science, and if it passes our peer review process (which determines whether it deserves to join the scientific literature) then we will publish it. In this way, no author

should ever receive the message that their article is scientifically sound but “not interesting enough” for our journal, or that their article is “only suited to a specialized audience”. As a result, we short circuit the vicious cycle of submit to a “top tier” journal; get reviewed; get rejected; submit to the next journal down the list; repeat until accepted and we are therefore able to place good science into the public domain as promptly as possible, with the minimum of burden on the academic community.

The most recent example of the way in which we separate pre-publication activity from post-publication activity is with the development of our Article-Level Metrics program. Article-level metrics start from the assumption that the best way to measure the worth of an article is to look at the actual article itself (and not the journal it happens to have been published in). Following from this, it seems obvious that the best way to evaluate any article is to make use of the collective opinion of all experts in the field (and not a small number of peer reviewers, or a small number of people who ultimately go on to cite the article). An evaluation of this type (which requires that people actually read, and then act in a variety of ways, on the article) can only be done after the article is published (and not before, as happens when a journal rejects a paper because it thinks it is not “impactful” enough). Therefore, our development of new tools to facilitate the post-publication evaluation of individual articles is a great example of us separating out what is most appropriately conducted pre-publication vs post-publication.

I have gone into some detail on these issues in a recent peer reviewed paper – PLoS One: background, future development, and article-level metrics and I would also recommend Shirley Wu’s excellent recent blog post on this topic.

2. Can you describe what Ambra/Topaz is and does?

Basically Ambra is our publishing platform, which runs on top of the Topaz infrastructure. For the full technical detail, the best information is found [here](#).

3. Can you talk a little bit more about the post-publication features of Ambra/Topaz?

Once an article is published, our platform allows users to leave feedback directly on the article. We were among the first journals to allow this, and we remain somewhat unique in this respect, although the concept is gaining broader acceptance and similar functionality is starting to appear at other publishers sites. Specifically, we allow users to leave a Note inline with a specific selection of text; or they can leave a general Comment on the entire article; or they can Star rate the article (on a 5 point scale in 3 categories). Comments and Notes form discussion threads, and users can then engage in debate on the points raised.

Users may not be anonymous, they must follow our guidelines for civilized academic debate, and when leaving feedback they are asked to declare any competing interests. I want to make it clear that the post-publication functionality

that we provide is not intended to provide post-publication peer review – it is post-publication discourse and feedback. All PLoS titles now share this same functionality.

4. Are other publishers besides PLoS using the Ambra/Topaz platform?

None that we are aware of although the NIH has an internal project. But Topaz is open source, so please contact us if you want to use it!

5. What article-level metrics does PLoS ONE provide?

As of today (August 2009), on every article, in every PLoS title we provide:

- Number of citations (as measured by Scopus and PubMedCentral)
- Number of social bookmarks (as recorded by CiteULike and Connotea)
- Number of star ratings left by users on our system
- Notes, Comments and any replies, as left by users of our system
- Number of blog posts written about the article (as counted by the blog aggregators Postgenomic, Nature Blogs and Bloglines)
- Specific trackbacks to the article from any web page using our trackback protocol

In September we will be adding additional citation data as measured by CrossRef and, most significantly, the online usage for each article (going back to the original publication date and reported on a monthly basis, broken down by HTML pageviews, PDF downloads, and XML downloads). This development in particular is very exciting – no other publisher has made this data available for such a large corpus of articles.

After this, the next data source we will add will be blog coverage as aggregated by ResearchBlogging.org, and in subsequent months we will be adding other metrics as and when we can identify high quality sources which meet our criteria.

Article-level metrics are a major development for PLoS and we believe that we are unique in the publishing industry with the transparent provision of such a range of article-level metrics. No other publisher provides as much (or any, in most cases) article-level data in such a comprehensive and open manner. It is our belief that once we have demonstrated what is possible, as well as the power of these metrics, the academic community will quickly begin to expect, and demand, this level of information from all journals. As a result the very nature of research reporting and evaluation will be improved as a result.

6. How do article-level metrics fit in with how PLoS ONE conducts peer review

We peer review all submissions for their scientific content, but we do NOT peer review them in order to determine whether they are high or low impact (or

interest or significance or relevance etc). Therefore, from the reader’s point of view, when you encounter a PLoS ONE article you do not necessarily know how impactful, or interesting, or significant, or relevant that article might be (without actually reading it!). In the traditional model, you would have some indication as to the likely importance of an article by a knowledge of the journal in which it was published in (although we argue that this way of determining quality is actually one of the worst methods you could use), however in PLoS ONE all you know is that the article is scientifically and methodologically sound (which are the only questions that our peer review process asks). Therefore, article-level metrics provide the reader with an indication as to the worth of an article once it is published. Until today, people have effectively said: this article was published in journal X, therefore knowing this one fact, I now know that the article is excellent/good/average/poor. I think that any sane person who considered that statement would realize how unscientific it was. With the advent of article-level metrics, a reader can now say “this article was published as part of the scientific literature, it is irrelevant which journal it was published in as I have now been given a variety of information about the article itself which will help me decide whether the article is excellent / good / average / poor for my own purposes”.

Therefore, article-level metrics do not supplant peer review and they also do not represent post-publication peer review. However they do provide the reader with new and valuable ways to do the post-publication evaluation and filtering of journal content.

7. What are PLoS ONE subject areas or portals?

We actually have several ways to ‘parse’ our content by topic: All content is assigned to one or more of our 52 topic areas (for example Pathology or Oncology). These topics are assigned by the authors themselves and an article can appear in more than one topic. Having made that classification, readers can then browse the topics or subscribe to an RSS feed per topic.

However, we appreciate that this is not a very flexible way to find content that doesn’t easily fall under our existing taxonomy structure. Therefore, we also have the ability to aggregate our articles into Collections. A Collection is literally just an aggregation tool (post-publication) – articles are still published as part of the normal run of the journal, but can then be assigned to join a Collection where they will also appear as part of a collection of related articles. For example, right now we have a very popular Paleontology Collection. As we publish new Paleontology articles they get added to this Collection to form a single location for all relevant articles in the field. This Collection functionality can also be used for the output of a single research effort, and an example of this is our Stress-Induced Depression and Comorbidities Collection which effectively replicated a “Special Issue” of a journal, and contained all the articles we published as written by the EUMOOD Research Consortia. Collections can be static, or can build up over time, and articles can appear in multiple Collections – as such they represent a very flexible way to re-present our content.

Then we have PLoS Hubs, which are under development right now. We see a Hub as a way to aggregate journal articles (along with other types of content) about a given topic into a single location. Once aggregated, we can then provide various community specific tools and services around this content. A Hub should not be thought of as a portal or an overlay journal – the distinguishing feature will be that a Hub will physically contain (and not just link out to) as much content as possible. Clearly, the easiest way to achieve this is by making the content Open Access, and so we also see Hubs as an opportunity to demonstrate the power of an Open Access copyright license. At the moment there is only one Hub (the PLoS Clinical Trials Hub) but this is a rather old implementation of the concept and only contains PLoS content – therefore we are proactively working on a new release which will include more of the functionality described above.

Future developments to our platform will involve the ability to tag articles (perhaps by some combination of curated and user generated tags) which will provide yet another way to dynamically aggregate the content.

8. How has the Ambra/Topaz platform handled the enormous growth of PLoS ONE?

Great! We had a few architecture issues in the past due to the bleeding edge nature of the platform, but all seven of our journals have now migrated to the same platform and no substantial issues have come up since the we migrated our Community Journals (back in early 2008).

9. What are your responsibilities at PLoS?

I am the Managing Editor of PLoS ONE (one of seven titles at PLoS). Although this position is an Editorial one, it is the position which is ultimately responsible for everything associated with the journal. By this I mean that although other departments may not report into me, I am ultimately responsible for the marketing, production, operations, web etc of the journal. If we have a problem with any aspect of the journal, it is me that makes sure it gets solved!

10. What did you do before starting to work at PLoS?

Well, I am a physicist from way back – I have a first degree in Physics with Astrophysics and a PhD in Underwater Optical Holography (which I always tell people sounds a lot more interesting than it was!). After my PhD I realized I wasn't interested in a life in academia and so I moved into Academic Publishing, which allowed me to stay in touch with academia and also to interact with leading researchers conducting the latest research. I started out at Institute of Physics Publishing, in Bristol UK, doing book acquisitions, then I moved to Holland to work for Kluwer Academic Publishers (KAP) to start up a reference work program for them. Via a series of moves I ended up running the KAP Earth, Environmental and Plant Sciences division – a large portfolio of books,

reference works and journals in those areas. Around the time that KAP merged with Springer I moved into Business Development for a year or so, which was an interesting period working on Springer's Open Choice program, online reference works and journal acquisitions among others. Then I left Springer, and also Holland, to move to California (my wife is from San Diego) where I worked for SAGE Publications, just North of LA, for 3 years. There I ran the SAGE US Journals division, which was made up of some 200 or so journals, mostly in the Social Sciences. In that position a large part of the job involved bidding on society titles, to publish them under contract. And then, finally, in March 2008 I moved to San Francisco, to work for PLoS and run PLoS ONE.

11. Do you want to talk about future plans for Ambra/Topaz?

The list of upcoming projects includes:

- More article-level metrics development
- RDFa implementation
- Automatic article relationships
- Semantic enhancement
- REST-based API
- The ingest and publication of many types of content / data (structured and unstructured)
- Tags
- Enhanced search and browse functionality
- A new process to submit articles directly to PubMed Central and other external repositories
- Direct access to our underlying triple store (sparql endpoint, RDFa)