Moving article-level metrics forward

Martin Fenner, Gobbledygook

November 11, 2009

In September PLoS started to show usage data (downloads, citations, but also use of social bookmarking services and blog posts) for all their published papers (article-level metrics at PLoS: addition of usage data). PLoS is not the first publisher to do that, but certainly the largest to date. Two Nature Network bloggers wrote about these changes back in June (The Scientist: On article-level metrics and other animals) and August (Gobbledygook: PLoS One: Interview with Peter Binfield), and a number of blogs commented on this new feature, including:

- A Blog around the Clock: Article-Level Metrics at PLoS – Download Data
- The Scholarly Kitchen: PLoS Releases Article-level Metrics
- BMJ Group Blogs: Richard Smith: The beginning of the end for impact factors and journals
- Blue Lab Coats: Article Level Metrics Debut at PLOS

There are a number of reasons why article-level metrics are a good idea, and I hope that many other journal publishers will follow. But in this blog post I want to talk about some of the shortcomings of the current implementation of article-level metrics.

Article-level metrics should be combined from different places

Full-text articles live in more than one place. Obviously at the journal publisher's website, but in many cases also in one or more institutional repositories and at PubMed Central (or similar places for papers not published in the life sciences). Which of these places produces the most reliable article-level metrics or should the HTML views, PDF downloads, etc. from all these places be combined? The decentralized nature of institutional repositories makes it especially difficult to combine usage statistics from them, but there are projects that try to tackle this problem. A unique identifier is required to combine the usage data from these different sources, and we have the DOI for that. PubMed Central and similar large repositories could not only start to provide their own usage data, but also combine them with the usage data from those journal publishers that already provide them.

Article-level metrics need author identifiers

Evaluating the "impact" of a researcher is one obvious use for article-level metrics. In order to be able to do that for more than a handful of researchers, we need unique author identifiers. This year we have had many discussions about author identifiers (including this blog and at the Science Online London Conference), and I hope that in 2010 we will finally see an evolving standard that is picked up by journal publishers. It would be in the interest of PLoS to combine their article-level metrics with an author identifier as soon as possible, most likely the proposed CrossRef ContributorID, rather than the Elsevier Scopus Author Identifier or the Thomson Reuters Researcher ID.

Article-level metrics should enhance literature searches

We all know how Google became the most popular search engine (Pagerank). And article usage data would be a tremendous boost for scientific literature databases such as PubMed. A literature search should sort the results by usage data (e.g. a combination of number of citations, HTML views and PDF downloads) rand not the rather boring publication date, author or journal name. Normally I would think that Google Scholar would be the first place to implement such a functionality, but I haven't seen much innovation from Google Scholar lately.

Article-level metrics should not only be numbers

As we don't want to reduce a paper to simple numbers, it is important to provide more than HTML views and PDF download counts. Citations counts are useful numbers, but linking to the citing papers is even more interesting. Similarly we want to see links to Faculty of 1000 recommendations and blog posts aggregated at ResearchBlogging.org. If we extend this further, we should probably start to think about a better name for article-level metrics. And I hope we never start to call this ALM.