



Novel Research Impact Indicators

Martin Fenner

Hannover Medical School, Hannover, Germany and Public Library of Science, San Francisco, CA, USA

mfenner@plos.org

Jennifer Lin

Public Library of Science, San Francisco, CA, USA

jlin@plos.org

Abstract

Citation counts and more recently usage statistics provide valuable information about the attention and research impact associated with scholarly publications. The open access publisher Public Library of Science (PLOS) has pioneered the concept of article-level metrics, where these metrics are collected on a per article and not a per journal basis and are complemented by real-time data from the social web or altmetrics: blog posts, social bookmarks, social media and other.

Key Words: article-level metrics; altmetrics; research impact

1. Introduction

Librarians have long since been applying usage counts for journal articles as a measure of value in order to maximize their acquisitions budgets. But the act of taking a journal volume from a library shelf left no traces twenty years

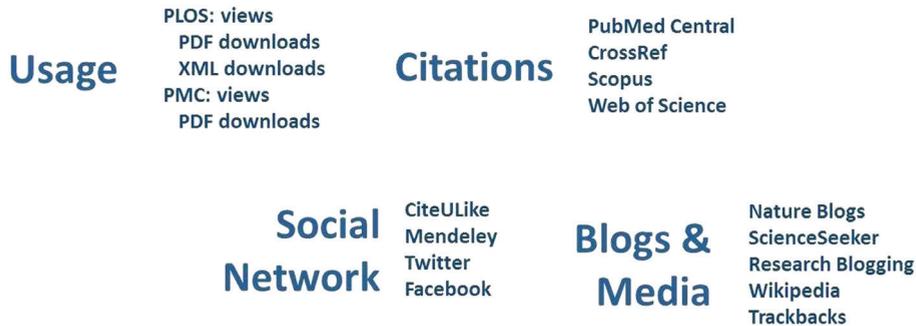
ago. As such, librarians have relied on COUNTER-compliant usage reports for the data they need to make pivotal decisions regarding which subset of published research they were going to make available to their community. Such data have offered a glimpse into the scope of usage for each journal under consideration.

The research community at-large has maintained a similar framework in assessing scholarly communications, whatever their needs might be. Across the research ecosystem, the prevailing paradigm of measuring the importance of research is based on the journal as a whole. However, a leading group of bibliometricians, research scientists, publishers, and policy-makers have begun to develop innovations in research assessment using alternative approaches such as article-level measurements of activity.

In today's digital environment, researchers have a broad spectrum of ways to access, manage and organize, share, comment, and cite others' research. This very environment has also enabled the collection and analysis of such activities. PLOS became the first publisher in 2009 (Neylon & Wu, 2009) to actively present measures of impact at the article level, providing a more focused and granular understanding of the importance and reach of a piece of research. With Article-Level Metrics (ALM) (PLOS, n.d.), PLOS displays transparent and comprehensive information about the usage and reach of published articles on the articles' pages so that each community can assess their value based on what is relevant and important to their different needs. ALMs offer direct, first-hand views of the dissemination and reach of research articles through an ever-increasing range of different metrics, from the more traditional usage and citation data to a mosaic of social media data. This multi-dimensional suite of indicators captures the research footprint from the moment of publication and dynamically tracks its impact over time. The existing suite of metrics is summarized in Figure 1.

Now that almost all articles are available online, it is possible to measure how an article is being read and regarded. By gaining a view into the manner in which researchers engage the research, we can start to determine the type of audience and the purpose of its dissemination. Different types of Web services, such as social bookmarking, social reference management, social news/recommendations, publisher-hosted comment spaces, and data repositories, all provide insights into the various manners of engagement with research that are possible.

Fig. 1: Article-Level Metrics collected and provided by PLOS.



New media and crowdsourcing tools form post-publication activities that assist researchers in the assessment of research articles on their own merits and they establish scholars' authority, augment peer review, broaden the scope of existing measures of impact that are tied to journal or publisher brand, and provide ways of discovering and filtering articles.

Furthermore, ALMs are a transparent way of mapping and analysing personal relationships between scientists, making them more quantifiable than ever before and allowing researchers to estimate which scholarly articles and journals are truly central to their individual flow of information.

ALMs can provide meaningful data that have the potential to become transformative, actionable information across multiple domains in research:

Research assessment

- Evaluation based on merits of actual research instead of that of the publishing journal
- In-depth, informed perspective for decision-making (funding, promotion, research ingestion, etc.) supported by transparent, comprehensive measures of impact

Research navigation

- Personalized literature search (navigate, filter, and sort) for focused research discovery

- Enhanced research discovery with valuable recommendations based on collective intelligence indicators

Research monitoring and tracking

- Efficient, streamlined way to stay informed of recent publications in a specific field and at large
- Survey of latest research trends based on most current metrics of article impact

Research process

- Up-to-date view of research progress, which can be easily shared (e.g., to institutional administrators, funders, etc.)
- Enhanced project design and implementation with an enhanced and precise view of research developments in any field
- Informed selection of collaborators based upon the impact of their work and relevance to yours

2. Article-Level Metrics in detail

The core value of ALMs is to increase the diversity of what we measure. This reflects our growing understanding of the need to measure impact beyond the academic community. At the same time, it is important for ALMs to include traditional measures. Citations remain a high-quality metric of impact on other researchers. Alongside citations, there are new measures that also tell us about forms of use that were largely invisible until recently and are able to do so with regularly updated context to the article far in advance of the period in which citations begin to accrue (which is often years after the article's publication date). This diverse set of impact indicators offers numerous ways to assess and navigate research most relevant to the field itself, including: usage, citations, social bookmarking and dissemination activity, media and blog coverage, and discussion activity and ratings.

The PLOS ALM suite includes citations from various sources: CrossRef, Web of Science, Scopus and PubMed Central. Google Scholar citation counts are not included because they cannot be automatically retrieved via an API.

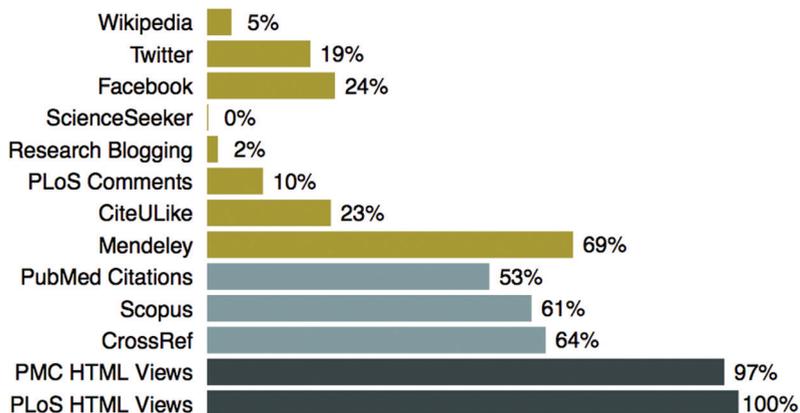
More than 90% of articles older than two years are cited at least once in any of these services, for all articles combined this number is about 60% (Figure 2).

Usage data are provided separately for HTML page views and PDF downloads, and from two separate sources: the PLOS website and the PubMed Central repository. There is no abstract landing page at the PLOS website. The majority of usage happens at the PLOS website: 83.6% of all HTML page views and 68.6% of all PDF downloads as of April 2013. Usage data from institutional repositories are currently not included into the PLOS Article-Level Metrics. PLOS and PubMed Central are not collecting any geolocation information for usage.

Altmetrics track the impact of a scholarly article on the social web. A metric can only give useful information about an article if a) it is available via an API using the DOI or other persistent identifier for the article, b) it tracks a sizeable portion of all articles, and c) it measures something of scholarly interest. The main limitation is a), and this, e.g., makes it a great challenge to track the news coverage of a scholarly article. The coverage by altmetrics sources varies widely between less than 1% and close to 70% (Figure 2).

The popularity of altmetrics sources also changes over time as scholarly communication patterns change. Whereas science blogging (ScienceSeeker and

Fig. 2: Proportion of articles covered by source. Metrics for 77,385 PLOS articles. Data collected April 11, 2013. Colour indicates ALM category (yellow = altmetrics, light blue = citations, dark blue = usage). Web of Science not shown because of license restrictions.

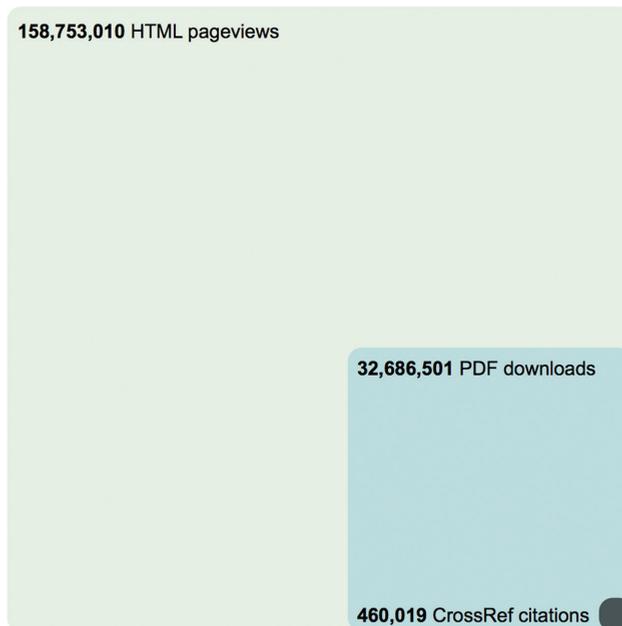


ResearchBlogging) and comments on the PLOS website have decreased relative to the number of articles published, Twitter has become more popular and we find tweets to 45% of all PLOS papers published since June 2012. Altmetrics is a diverse group of metrics and it is helpful to group them into subgroups of related services: academic bookmarking tools (Mendeley, CiteULike), social shares (Twitter, Facebook), and blogs and media (ScienceSeeker, Research Blogging, Wikipedia).

The overall story that the metrics tell is a more comprehensive one than the sum of its pieces. As of April 2013, all PLOS papers combined received 158 million pageviews and were downloaded 32 million times (Figure 3). Only 460K, or 0.3% of the HTML pageviews, resulted in a citation, indicating that a focus on citation metrics alone would miss more than 99% of the activity around a paper.

While each metric contributes to the evolving story of science, they offer a different piece of the summary. This is to say that no single indicator alone

Fig. 3: Article-Level Metrics for 77,385 PLOS papers published until April 11, 2013. HTML page views and PDF downloads from PLOS journals' website.



can represent research impact: page views provide a strong signal of interest. Social bookmarking sites, particularly those that are focused on researchers such as Mendeley and CiteuLike provide information on what researchers are collecting into their personal libraries — a strong signal of relevance and interest. In combination with these measures, wider social media activity can be highly informative. For example, Twitter can provide rich information on who is interacting with what articles, and why. Each of the metrics offers a view into the conversation surrounding the research, though each in its own way. And as all ALMs are based on web services, we can cross-validate different types of metrics against each other, providing a useful set of checks and balances.

3. A case study

To better demonstrate the diversity of article-level metrics, we analysed a sample set made up of all PLOS articles published by authors from a single institution (Hannover Medical School, the affiliation of the first author of this paper). One hundred eighty-nine articles were found by a free-text search for the affiliation “Hannover Medical School” on April 11, 2013, using the PLOS Search API (Fenner, 2012). The dataset and R script to collect the data is available for download (Fenner & Lin, 2013). The free-text search for the affiliation could have missed some articles, as affiliation names are not reported consistently (e.g., the variants “Hanover Medical School” or “Medizinische Hochschule Hannover”).

Similar to almost all PLOS articles, this set of articles shows a very strong correlation of 4:1 between the number of HTML page views and PDF downloads (Figure 4). This correlation is independent of article age (not shown) or journal. Two of the three most viewed papers are also the top-cited papers in this set, but overall there is only a weak correlation between usage and citations, consistent other PLOS articles.

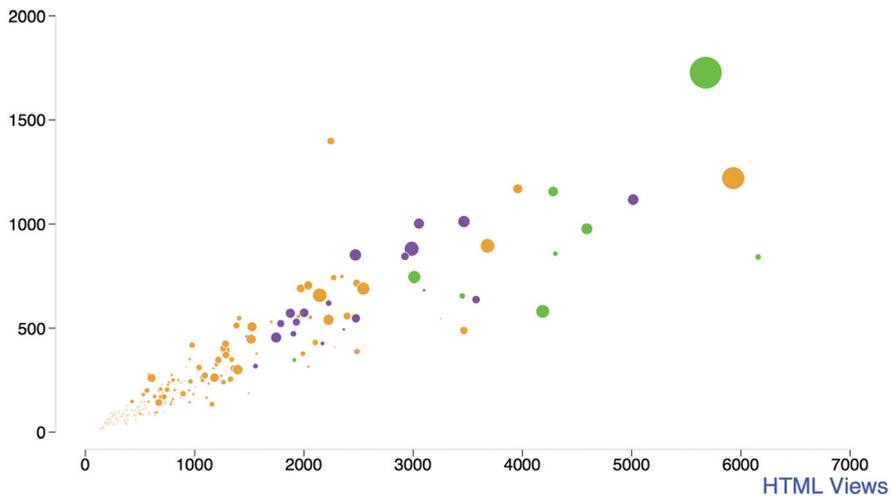
One interesting outlier is an article with a PDF/HTML ratio of 0.53 (Walter et al., 2009) (the orange bubble with almost 1,500 PDF downloads in Figure 4). To better understand this pattern, we looked at HTML page views and PDF downloads over time (Figure 5). The temporal usage pattern is typical for most journal articles with the majority of downloads in the first few months

Fig. 4: HTML views vs. PDF downloads for 189 PLOS articles published by Hannover Medical School authors. Colours correspond to PLOS journals (PLOS ONE = green, PLOS Medicine, PLOS Pathogens, PLOS Neglected Tropical Diseases = purple, PLOS Biology, PLOS Genetics, PLOS Computational Biology = green).

HTML views vs. PDF downloads

Data for PLOS website for 189 PLOS articles with authors affiliated with Hannover Medical School. Bubble size correlates with number of CrossRef citations, and color with PLOS journal. Data collected April 11, 2013.

PDF Downloads



after publication. During this time, a page view almost always resulted in a PDF download, resulting in a very high PDF/HTML ratio. Starting with month 6 after publication, the PDF/HTML ratio dropped to much lower numbers.

Many researchers use a reference manager to organize PDF files of scholarly articles downloaded to their computers. Mendeley is one of the more popular reference managers and provides the number of users that bookmarked an article. In our sample of 189 articles we see a good correlation between the number of PDF downloads and the number of Mendeley bookmarks (Figure 6). There are a small number of outliers, e.g., the most-bookmarked article[6] has only an average number of PDF downloads.

This case study only scratched the surface of what Article-Level Metrics can do, but it clearly demonstrates the value of collecting metrics at the article

Fig. 5: Monthly HTML views and PDF downloads at the PLOS website for (Jessen et al., 2011). Data collected April 15, 2013.

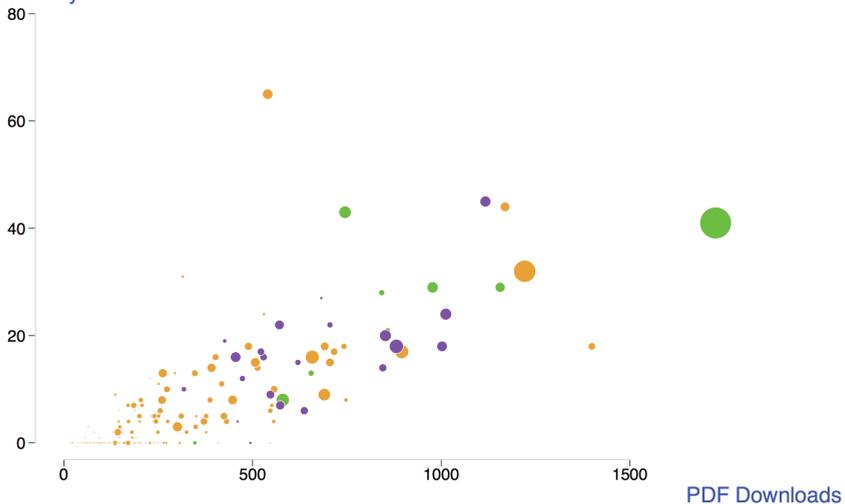


Fig. 6: PDF downloads vs. Mendeley bookmarks for 189 PLOS articles published by Hannover Medical School authors. Colours correspond to PLOS journals (PLOS ONE = green, PLOS Medicine, PLOS Pathogens, PLOS Neglected Tropical Diseases = purple, PLOS Biology, PLOS Genetics, PLOS Computational Biology = green).

PDF downloads vs. Mendeley bookmarks

Data for PLOS website for 189 PLOS articles with authors affiliated with Hannover Medical School. Bubble size correlates with number of CrossRef citations, and color with PLOS journal. Data collected April 11, 2013.

Mendeley Bookmarks



level, of including metrics other than citations, and of looking at metrics not as single numbers, but in the context to each other and over time.

4. Conclusions

Article-Level Metrics have clearly opened the door for novel approaches to research impact assessment. At present their value is primarily in research navigation and research monitoring. But this is a highly dynamic field that is moving towards wider use and standardization, and it will probably not be too long before we see Article-Level Metrics routinely used to aid in research assessment. At the same time ALM are an excellent toolset for studying the scholarly communication process itself, and we can learn a great deal about how research is disseminated, discussed and reused after publication.

References

- Neylon, C., & Wu, S. (2009). Article-level metrics and the evolution of scientific impact. *PLOS Biology* 7(11), e1000242. doi:10.1371/journal.pbio.1000242. Retrieved May 11, 2013, from <http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1000242>.
- PLOS (n.d.). *PLOS Article-level metrics (ALMs): measuring the impact of research*. Retrieved April 15, 2013, from <http://article-level-metrics.plos.org>.
- Fenner, M. (2012, July 20). *Example visualizations using the PLOS Search and ALM APIs*. Retrieved April 11, 2013, from <http://api.plos.org/2012/07/20/example-visualizations-using-the-plos-search-and-alm-apis/#more-1661>.
- Fenner, M., & Lin, J. (2013). *Article-Level Metrics Hannover Medical School*. doi:10.6084/M9.FIGSHARE.681737. Retrieved October 4, 2013, from http://figshare.com/articles/Article_Level_Metrics_Hannover_Medical_School/681737.
- Jessen, F., Wiese, B., Bickel, H., Eiffländer-Gorfer, S., Fuchs, A., Kaduszkiewics, H., et al. (2011). Prediction of dementia in primary care patients. *PLOS ONE* 6(2), e16852. doi:10.1371/journal.pone.0016852. Retrieved May 11, 2013, from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0016852>.
- Walter, H., von Kalckreuth, A., Schardt, D., Stephan, A., Goschke, T., & Erk, S. (2009). The temporal dynamics of voluntary emotion regulation. *PLOS ONE* 4(8), e6726. doi:10.1371/journal.pone.0006726. Retrieved May 11, 2013, from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0006726>.