

How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIs

Zohreh Zahedi¹, Martin Fenner² & Rodrigo Costas³

¹ z.zahedi.2@cwts.leidenuniv.nl; ² mfenner@plos.org; ³ rcostas@cwts.leidenuniv.nl

^{1,3} Centre for Science & Technology Studies (CWTS), Leiden, The Netherlands

² Public Library of Science (PLOS), San Francisco, US

Introduction

Altmetrics track the impact of scholarly works on the social web. The term was introduced in 2010 (Priem, et al.) as an alternative way of measuring the broader research impact of scholarly outputs using the social web; aimed at enhancing and complementing the more traditional ways of impact assessment via citations. The initial phase of altmetrics has been characterized by the development of diversity of tools that aim to track ‘real-time’ impact of scientific outputs (Wouters & Costas, 2012). Several studies have started to analyze the presence of altmetrics across scientific publications (Priem, Piwowar, & Hemminger, 2012; Zahedi, Costas & Wouters, 2014; Costas, Zahedi, & Wouters, 2014; Thelwall et al., 2013). However, little is still known about the quality of altmetric data obtained by these providers. It seems that similar metrics differ across different providers due to the difference in collection time, data sources and methods of collection among altmetrics providers (Chamberlain, 2013). Hence, the assessment of the quality, reliability and consistency of altmetric data is crucial in order to be able to introduce altmetrics for research assessment purposes. This study targets to investigate 3 main altmetrics providers (PLOS ALM, Altmetric.com and Mendeley) and to test the accuracy and quality of their metrics for a same set of publications. The research questions are as follows:

1. Are there differences across these three altmetrics providers in the metrics for the same set of publications?
2. If there are differences, what are possible factors that explain these differences?

Data and Methodology

This study is based on all PLOS ONE publications from 2013 (31,408 articles), retrieved from the full PLOS ALM Dataset on 14 Jan 2014¹. A random sample of 1,000 publications from this data set has been extracted. DOIs were used for collecting the metrics automatically from three providers of altmetrics data: PLOS ALM, Altmetric.com and Mendeley using their REST APIs. The data collection was performed at the same date and time (11 AM CET on February 11, 2014). The R statistical analysis software version 3.0.2 and the rOpenSci alm package² were used to obtain the data from the PLOS ALM REST API v3³, and to generate a CSV report. For gathering the altmetric data from the Mendeley and Altmetric.com, the responses provided on search requests using DOI's were downloaded per API search request separately in Java Script Object Notations (JSON) format on the basis of individual DOIs and parsed by using the additional JAVA library from within the SAS software⁴. Finally, the data transformed into a comma separated value format (CSV) and imported in SQL in order to join the files from the three altmetrics providers and to perform further analysis.

Results

Coverage of PLOS ONE publications across altmetrics providers

Table 1 shows the coverage of the 1000 PLOS ONE publications by these altmetrics providers. We have focused only on 3 altmetric indicators: Mendeley readerships, Twitter counts and Facebook counts. The number of publications with at least one metric (Mendeley readers, tweet and Facebook counts) show that Mendeley has the highest coverage, followed by PLOS ALM and Altmetric.com. There are more publications with at least one tweet in Altmetric.com vs. PLOS ALM and the other way around for Facebook counts (PLOS ALM has a higher coverage than Altmetric.com).

Table 1. coverage of PLOS ONE publications by altmetrics providers

Altmetrics Providers	No of pubs with at least one Mendeley reader	No of pubs with at least one tweet	No of pubs with at least one Facebook count
PLOS ALM	588	325	Comments 138 Likes 168 Shares 261 Total 261
Altmetric.com	460	490	210
Mendeley	900	--	--

¹ <http://article-level-metrics.plos.org/plos-alm-data/>

² <http://ropensci.org/packages/alm.html>

³ PLOS REST API v3 described in details here: <http://articlemetrics.github.io/docs/api/>

⁴ The additional functionality from the ‘proc groovy’ which is a java development environment added to SAS environment for parsing and reading the JSON format and returning the data as an object.

Proportion of PLOS ONE altmetrics across altmetrics providers

Table 2 shows the total counts for the same set of altmetrics for the sampled publications. Mendeley provides the highest counts of readerships compared to PLOS ALM and Altmetric.com⁵ (this value is as twice as and more than two times than the total value by Altmetric.com and PLOS ALM). Regarding the total number of tweets, Altmetric.com provides the highest number of tweets for this dataset (Altmetric.com counts for tweets is around 1.6 times higher than the values from PLOS ALM). The counts from Facebook also show important differences with PLOS ALM having much higher counts compared to Altmetric.com.

Table 2. Proportion of PLOS ONE altmetrics across altmetrics providers

Altmetrics Providers	Total no of Mendeley reader counts	Total no of Tweets	Total no of Facebook counts
PLOS ALM	2204	2484	Comments 2240 Shares 3877 Likes 4672 Total 10789
Altmetric.com	2734	4129 Post counts 3828 Unique users counts	615 Post Counts 582 Users counts
Mendeley	5330	--	--

Conclusion and Discussions

Data quality and consistency across different altmetrics providers is an important issue. Therefore it is important to know how and why similar metrics differ across different providers. The findings showed that although these three studied altmetrics providers share some data sources (i.e. Facebook, Twitter and Mendeley) and also the date and time of altmetrics data collection from these three providers have been controlled in this study, altmetrics data reported for the same dataset of publications is not consistent among them and large differences have been observed. Therefore, in order to know how consistent the providers are, we need to know how these providers collect different metrics. Mendeley collects readership counts for documents saved by users in their own libraries using clustering algorithm which run daily across the entire Mendeley catalog. The readership number reported is the size of the document cluster (Gunn, 2013). Altmetric.com tracks articles through RSS feeds. They add a specific news-outlets or blogs to their database in which they automatically track the RSS feeds. When there is a publication id mentioned (e.g. arXiv ids or PubMed ids), they track the respective mentions to the articles. However, Mendeley counts aren't updated in real time in Altmetric.com and there can be lag of up to a week between the Mendeley readers count reported by Altmetric.com and the counts on Mendeley.com⁶. In PLOS, Mendeley readership counts are collected from the Mendeley API using the Mendeley UUID as unique identifier. For Facebook likes, shares, comments and total counts, PLOS use the Facebook link_stat API⁷ using the canonical URL of the article on the PLOS Journals platform (Fenner & Lin, 2014). For twitter, PLOS collects the tweets about articles by querying the Twitter streaming API for links to articles on the PLOS journals platform using an in-house Java application. All PLOS ALM API data are refreshed daily, the first 30 days after publication of the article, four times a month in the first year after publication and then monthly (Lin & Fenner, 2013).

Apparently all the three providers (except for Mendeley counts by Altmetric.com) collect the metrics in real time; therefore, there might be some differences in their methodology that could explain the divergence between them (e.g. the use of different APIs for Facebook and Twitter counts). Altmetric.com counts only public Facebook wall posts, whereas PLOS ALM collects all posts. For Mendeley readerships, the PLOS ALM service stores the UUID permanently and uses that identifier to collect information from Mendeley, differences and changes in the UUID in Mendeley could probably explain the differences (plus possible time lags in the data collection). As a consequence, PLOS has switched to not store Mendeley UUIDs permanently in March 2014.

Much needs to be done to improve the consistency of altmetrics data across different providers.

REFERENCES

- Chamberlain, S. (2013). Consuming Article-Level Metrics: Observations and Lessons. *Information Standards Quarterly*, 25(2), 4–13.
- Costas, R., Zahedi, Z., & Wouters, P. (2014). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*. DOI: 10.1002/asi.23309.
- Fenner, M., Lin, J. (2014). The Mystery of the Missing ALM, PLOS Blog, 6th March, 2014. Retrieved from: <http://blogs.plos.org/tech/facebook/>
- Gunn, W. (2013). Social Signals Reflect Academic Impact: What it Means When a Scholar Adds a Paper to Mendeley. *Information Standards Quarterly*, 25(2): 33-39.
- Lin, J., & Fenner, M. (2013). Altmetrics in Evolution: Defining & Redefining the Ontology of Article-Level Metrics. *Information Standards Quarterly* 25(2): 20-26.
- Priem, J., Taraborelli, D., Groth, P., and Neylon, C. (2010). Altmetrics: a manifesto. Retrieved from: <http://altmetrics.org/manifesto/>
- Priem, J., Piwowar, H., & Hemminger, B. H. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. ArXiv: 1203.4745v1
- Thelwall, M., Haustein, S., Larivière, V., and Sugimoto, C.R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE* 8(5): e64841.

⁵ According to Altmetric.com unique_users_count refers to the count of those users and posts_count shows the number of unique items created by those users that mention this article. See: <http://support.altmetric.com/knowledgebase/articles/95165-sample-api-response>

⁶ <http://support.altmetric.com/knowledgebase/articles/95165-sample-api-response>

⁷ https://developers.facebook.com/docs/reference/fql/link_stat/

Wouters, P., Costas, R. (2012). Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century. Utrecht: SURF foundation. Retrieved from: www.surfoundation.nl/nl/publicaties/Documents/Users%20narcissism%20and%20control.pdf

Zahedi, Z., Costas, R. & Paul Wouters. (2014). How well developed are Altmetrics? Cross disciplinary analysis of the presence of 'alternative metrics' in scientific publications, 101(2): 1491-1513. DOI: 10.1007/s11192-014-1264-0