Consuming Article-Level Metrics: Observations and Lessons from Comparing Aggregator Provider Data

by Scott Chamberlain

Introduction

The Journal Impact Factor (JIF; owned and published by Thomson Reuters)[1],[2] is a summation of the impact of all articles in a journal based on citations. Publishers have used the JIF to gain recognition, authors are evaluated by their peers based on the JIF of the journals they have published,[3] and authors often choose where to publish based on the JIF. The JIF has significant flaws, including being subject to gaming[4] and not being reproducible.[5] In fact, the San Francisco Declaration on Research Assessment has a growing list of scientists and societies that would like to stop the use of the JIF in judging work of scientists.[6] An important critique of the JIF is that it doesn't measure the impact of individual articles—clearly not all articles in a journal are of the same caliber. Article-level metrics measure the impact of individual articles, including usage (e.g., pageviews, downloads), citations, and social metrics (or *altmetrics*, e.g., Twitter, Facebook).[7] Article-level metrics have many advantages over the JIF, including:

- **Openness** – Article-level metrics are largely based on data that is open to anyone (though there are some that aren't, e.g., Web of Science, Scopus). If data sources are open, conclusions based on article-level metrics can be verified by others and tools can be built on top of the article-level metrics.
- **Speed** – Article-level metrics are nearly real-time metrics of scholarly impact.[7] Citations can take years to accrue, but mentions and discussion that can be searched on the web take hours or days.
- **Diversity of sources** – Article-level metrics include far more than just citations and provide metrics in a variety of domains, including discussion by the media (mentions in the news), discussion by the public (Facebook likes, tweets), and importance to colleagues (citations).

There are many potential uses for article-level metrics, including:

- **Research** – As article-level metrics rise in use and popularity, research on article-level metrics themselves will inevitably become a more common use case. Some recent papers have answered the questions: How do different article-level metrics relate to one another?[8],[9] What is the role of Twitter in the lifecycle of a paper?[10] Can tweets predict citations?[11],[12] These questions involve collecting article-level metrics in bulk from article-level metrics providers and manipulating, visualizing, and analyzing the data. This use case often requires using scripting languages (e.g., Python, Ruby, R) to consume article-level metrics. Consuming article-level metrics from this perspective is somewhat different than the use case in which a user views article-level metrics hosted elsewhere in the cloud. This use case is the target use case with which this paper is concerned.
- **Credit** – Some scholars already put article-level metrics on their CVs, usually in the form of citations or JIF's. With the rise of article-level metrics, this will become much more common, especially with initiatives like that of the U.S. National Science Foundation (NSF) that now allows scholars to get credit for *products*, not just papers—and products like videos or presentations cannot be measured by citations or JIF's. This use case will involve

scholars with a wide variety of technical skills and will be made easy with tools from ImpactStory or other providers.[13]

- **Filtering** – Scholars cannot possibly find relevant papers efficiently given that there are now tens of thousands of scholarly journals. Individual article-level metrics components can be used to filter articles. For example, many scientists use Twitter and are more likely to view a paper that is tweeted often—in a way leveraging article-level metrics. Article-level metrics can also be used to filter more directly. For example, article-level metrics are now presented alongside papers, which can be used to make decisions about what papers to read and not to read. Readers may be drawn, for example, to a paper with a large number of tweets or blog mentions.

In this paper I discuss article-level metrics from the perspective of developing and using scripting interfaces for article-level metrics. From this perspective, there are a number of considerations: where can you get article-level metrics data, data consistency, data provenance, article-level metrics in context, and technical barriers to use.

Article-level metrics data providers

There are a number of publishers now presenting article-level metrics for peer-reviewed articles on their websites (for examples, see Wiley-Blackwell, Nature, Public Library of Science (PLOS), Frontiers, and Biomed Central). Most of these publishers do not provide public facing APIs (Application Programming Interface—a way for computers to talk to one another) for article-level metrics data, but instead use aggregators to provide article-level metrics data on their papers. One exception is PLOS, which collects their own article-level metrics and provides an open API to use their article-level metrics data. At the time of writing, there are four major entities that aggregate and provide article-level metrics data: PLOS,[14] ImpactStory,[15] Altmetric,[16] and Plum Analytics[17] (see Table 1 for details). There are a few other smaller scale article-level metrics providers, such as CitedIn[18] and ScienceCard,[19] but they are relatively small in scope and breadth. There are some similarities and differences among the four providers, which may help in deciding which service to use for a particular purpose (see also Table 3).

Table 1: Details on the four largest article-level metrics providers.

Variable	PLOS	ImpactStory	Altmetric	Plum Analytics
Open API?	Yes	Yes	Limited[c]	No
Data format	JSON,JSONP,XML	JSON	JSON,JSONP	JSON
Granularity[a]	D,M,Y	T	I	T
API Authentication	API key	API key	API key	API key
Business type	Publisher	Article-level metrics provider	Article-level metrics provider	Article-level metrics provider
For-profit	No	No	Yes	Yes
Income based on	Page charges	Publishers/Grants	Publishers	Institutions

Variable	PLOS	ImpactStory	Altmetric	Plum Analytics
Rate limiting	Not enforced	Not enforced[b]	1 call/sec. [c]	Unknown
Products covered	Articles	Many[d]	Many[e]	Many[f]
Software clients	R[g]	R,Javascript[h]	R,Python,Ruby,iOS[i]	Unknown

Notes:
[a] D: day; M: month; Y: year; T: total; I: incremental summaries
[b] Note: They recommend delaying a few seconds between requests
[c] Also hourly and daily limits enforced; using API key increases limits
[d] articles, code, software, presentations, datasets
[e] articles, datasets, books
[f] articles, code, software, presentations, datasets, books, theses, etc. (see http://www.plumanalytics.com/metrics.html for a full list)
[g] https://github.com/ropensci/alm
[h] R (https://github.com/ropensci/rimpactstory), Javascript (https://github.com/highwire/opensource-js-ImpactStory)
[i] R (https://github.com/ropensci/rAltmetric), Python (https://github.com/lnielsen-cern/python-altmetric), Ruby (https://github.com/ldodds/altmetric), iOS (https://github.com/shazino/SZNAltmetric)

The four providers overlap in some sources of article-level metrics they gather, but not all (see Table 3). The fact that the sources are somewhat complementary opens up the possibility that different metrics can be combined from across the different providers to get more a complete set of article-level metrics. For those that are complementary, this should be relatively easy, and we don't have to worry about data consistency. However, when they share data sources, one has to choose which data provider to use tweets from, for example; and data may not be consistent between providers for the same data source (see the *Consistency* section below).

One of the important aspects of article-level metrics is that most of the data is from article-level metrics aggregators like ImpactStory who aren't creating the data themselves, but rather are collecting the data from other sources that have their own licenses. Thus, data licenses for PLOS, ImpactStory, Altmetric, and Plum Analytics generally match those of the original data provider (e.g., some data providers do not let anyone cache their data).

Consistency

Now that there are multiple providers for article-level metrics data, data consistency is an important consideration. For example, PLOS, ImpactStory , Altmetric, and Plum Analytics collect article-level metrics from some of the same data sources. But are the numbers they present to users consistent for the same paper or are they different due to different collection dates, data sources, or methods of collection? Each of the aggregate article-level metrics providers may collect and present article-level metrics as relevant for their target audience. Thus, as article-level metrics consumers and researchers, we need to have a clear understanding of the potential pitfalls when using article-level metrics data for any purpose, especially research where data quality and consistency is essential.

For this study a set of 565 articles were used, identified using their DOIs, from PLOS journals only; this way all four providers would have data on the articles. Metrics were collected from each of the four providers for each of the 565 DOIs using as primary sources Citeulike, Scopus, PLOS-Counter (usage data: html, xml, and pdf views), PubMed Central (PMC), Facebook, Mendeley, and Twitter. (Data was excluded from Plum Analytics for Citeulike as it was not

provided, but they do collect it.[20] In addition, Facebook data was excluded from Plum Analytics results because it was unclear how to equate their data with the data from the other providers.) For each DOI, the maximum difference between values (i.e., providers) was calculated and the distribution was plotted for seven article-level metrics that were shared among the providers. Figure 1 shows that, at least with respect to absolute numbers, PMC citations are very different among providers, while PLOS views (html + pdf views, relevant only to PLOS ALM, ImpactStory, and Plum Analytics) are somewhat less variable among providers. The remaining metrics were not very different among providers, with most values at zero, or no difference at all.
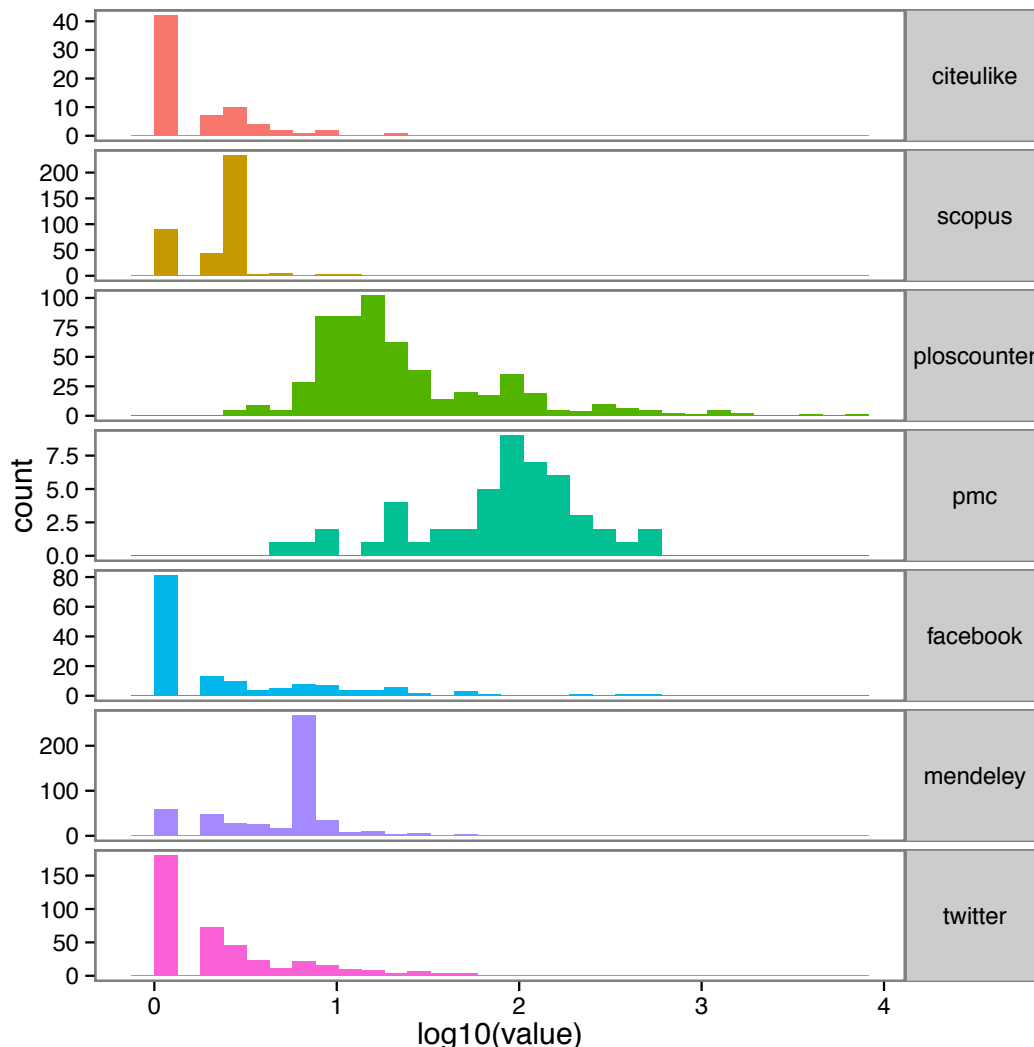

Figure 1: Distribution of absolute differences in least and greatest value of each of seven different article-level metrics

Note: Calculated on a set of 565 DOIs from Altmetric, ImpactStory, and PLOS ALM. Values were log10 transformed to improve visual comprehension. Metrics: citeulike = number of Citeulike bookmarks; scopus = number of citations; ploscounter = number of pdf views + html views; pmc = number of Pubmed Central full text + pdf views; facebook = number of Facebook shares; mendeley = number of Mendeley readers; twitter = number of tweets mentioning article.

What are some possible reasons why similar metrics differ across providers? First, data could be collected from different middle-men. For example, Twitter data is notorious for not being persistent. Thus, you either have to query the Twitter "firehose" constantly and store data, or go through a company like Topsy (which collects Twitter data and charge customers for access)

to collect tweets. Whereas ImpactStory collects tweets from Topsy, PLOS collects tweets from the Twitter firehose, and Altmetric collects tweets using a combination of the Twitter search and streaming APIs. Second, data could be collected at different times, which could easily result in different data even when collected from the same source. This is especially obvious as ImpactStory collects some metrics via the PLOS ALM API, so those metrics that ImpactStory has from the PLOS ALM API should be the same as those that PLOS has. Fortunately, date is supplied in the data returned by three of the providers (PLOS ALM, ImpactStory, and Altmetric). Thus, whether or not date could explain differences in metrics from the various providers was examined. Figure 2 shows that there are definitely some large differences in values that could be due to differences in the date the data was collected, but this is not always the case (i.e., there are a lot of large difference values with very small difference in dates).
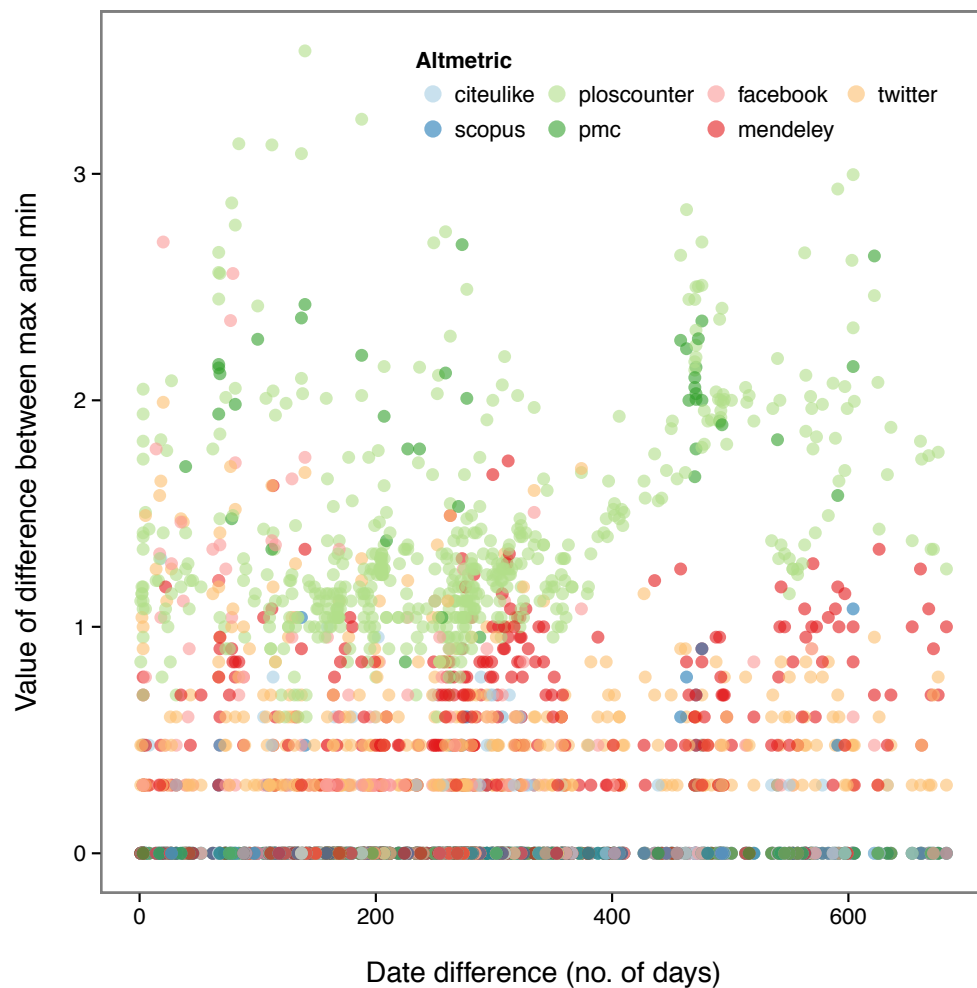

Figure 2: Distribution of absolute differences in least and greatest value of each of seven different article-level metrics

Note: Calculated on a set of 565 DOIs from Altmetric, ImpactStory, and PLOS ALM. Values were log10 transformed to improve visual comprehension. See Figure 1 for explanation of the specific article-level metrics.

The previous analyses were a rough overview of hundreds of DOIs. To determine the differences among providers in more detail, a set of 20 DOIs from the set of 565 were used. Figure 3 shows the value of each altmetric from each of the providers for each of the 20 DOIs.

Note that in some cases there is very close overlap in values for the same altmetric on the same DOI across providers, but in some cases the values are very different.
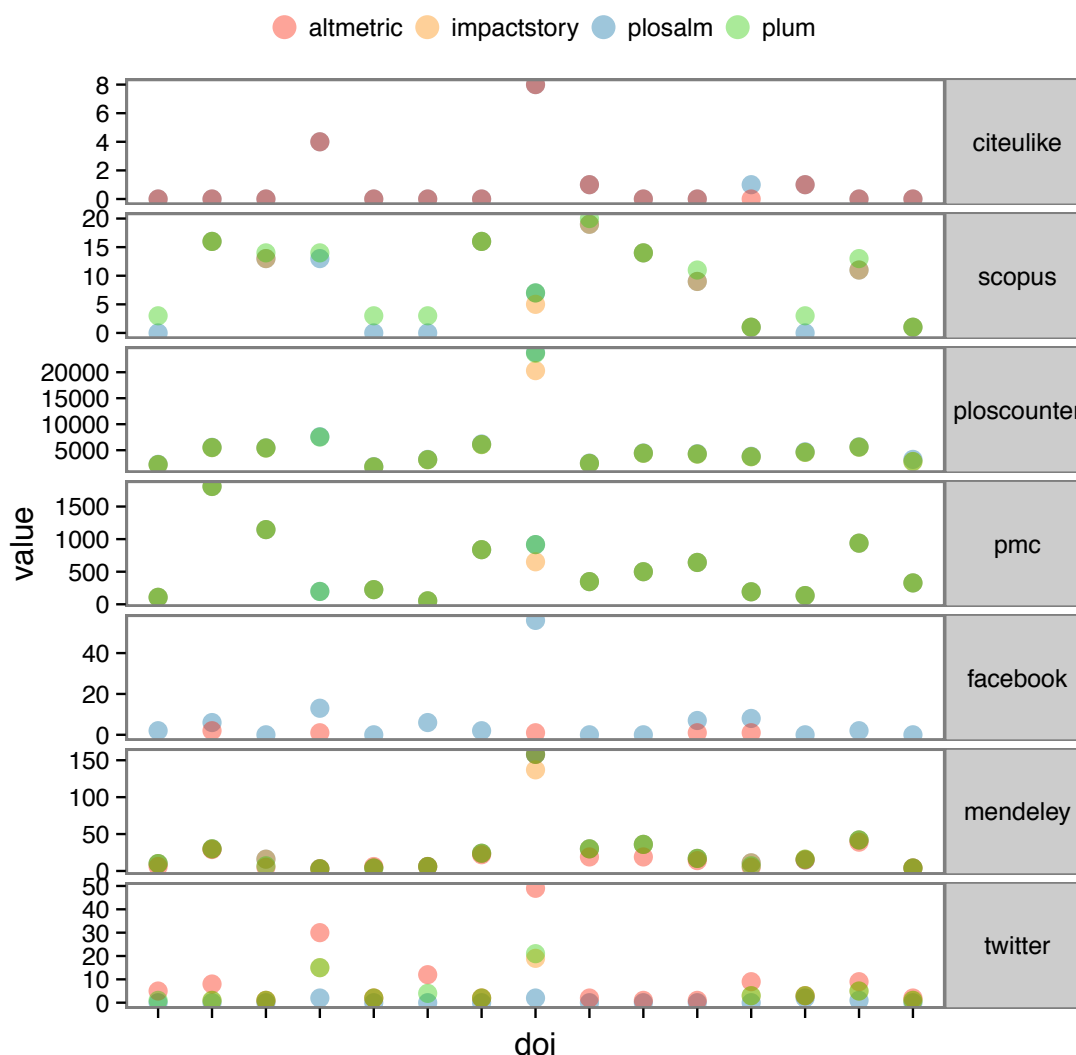

Figure 3: A comparison of seven different article-level metrics on a set of 20 DOIs from Altmetric, ImpactStory, and PLOS.

Note: This demonstrates how article-level metrics can be very similar across providers for some DOIs, but very dissimilar for others. See Figure 1 for explanation of the specific article-level metrics.

A particular example of these results may be instructive. Table 2 details the results of using the API of each the four providers to combine data from different sources for the DOI 10.1371/journal.pbio.1001118.[22] There are many metrics that have exactly the same values among providers, but there are also differences, which could be explained by the difference in the collection date. For example, PLOS ALM gives 3860 for combined PLOS views, while ImpactStory gives 3746 views. This is undoubtedly explained by the fact that PLOS ALM data was last updated on May 31, 2013, while ImpactStory's data was last updated on May 18, 2013. There are some oddities, however. For example, Altmetric gives nine tweets, ImpactStory and Plum Analytics only give three tweets, while PLOS ALM gives zero. ImpactStory's data was updated more recently (May 18, 2013) than that of Altmetric (July 28, 2012), which suggests

something different about the way tweets among the two providers are collected as updated date alone cannot explain the difference. In fact, ImpactStory collects tweets from Topsy, while Altmetric collects tweets with combination of Twitter search and streaming APIs, which leads to different data. Meanwhile, PLOS ALM collects tweet data from the Twitter firehose.

Table 2: Example of combining results across three data providers on one DOI.

Note: The dates that data were last modified are the same for PLOS ALM and Altmetric, but different for ImpactStory. Missing values represent data that is not given by that provider. See Figure 1 for explanation of specific article-level metrics.

Provider	citeulike	scopus	ploscounter	pmc	facebook	mendeley	twitter	Date modified
PLOS ALM	1	1	3860	192	8	11	0	2013-05-31
Altmetric	0				1	5	9	2012-07-28
ImpactStory		1	3746	192		11	3	2013-05-18
Plum Analytics		1	3746	192			3	unknown

The above findings on data consistency suggest that article-level metrics are inconsistent among providers of aggregate article-level metrics. Casual users, and especially those conducting article-level metrics research, should use caution when using article-level metrics data from different providers.

A crosswalk among providers

Each of the four providers, of course, has the right to collect metrics as needed for their purposes, but as article-level metrics consumers, we should be able to compare data from the same source across providers. When similar data sources are collected by article-level metrics providers, ideally, there should be a way to map data, e.g., from Twitter for PLOS, ImpactStory, Altmetric, and Plum Analytics. Table 3 provides a sample crosswalk of metrics for the same data source among providers.

Table 3: Crosswalk between article-level metrics data collected by the three data providers.

Note: These variables relate to one another across providers, but the data may be collected differently, so, for example, article-level metrics collected for Twitter may differ between PLOS, ImpactStory and Altmetric. Where data sources are shared among at least two providers, only those fields were used that would give the same data if collected on the same date and all other things being equal. For example, PLOS ALM's field *pubmed* is equivalent to ImpactStory's *pubmed:pmc_citations* field.

Data source	PLOS [a]	ImpactStory [b]	Altmetric [c]	Plum Analytics [d]
Biod	biod	No	No	No
Bloglines	bloglines	No	No	No
Nature blogs	nature	No	No	No
Researchblogging	researchblogging	No	No	ResearchBlogging
WebOfScience citations	webofscience	No	No	No
Dryad	No	dryad:total_downloads package_views	No	Views, downloads
Figshare	No	figshare:views shares downloads	No	Recommendations, downloads, views
Github	No	github:forks stars	No	Collaborators, downloads,

Data source	PLOS [a]	ImpactStory [b]	Altmetric [c]	Plum Analytics [d]
				followers, forks, watches, gists
PLOS Search	No	plossearch:mentions	No	No
Slideshare	No	slideshare:favorites views comments downloads	No	Downloads, favorites, comments
Google+	No	No	cited by gplus count	No. +1's
MSM	No	No	cited by msm count	No
News articles	No	No	Yes	Yes
Reddit	No	No	cited by rdts count	Comments, upvotes-downvotes
Citeulike	citeulike	citeulike:bookmarks	No	Citeulike
Crossref	crossref	plosalm:crossref [e]	No	No
PLOS ALM	counter(pdf_views + html_views)	plosalm(html_views, pdf_views) [f]	No	Views of abstract, figures, full text, html, pdf, supporting data
PMC	pmc	plosalm:pmc_full-text + pmc_pdf [g]	No	No
PubMed	pubmed	pubmed:pmc_citations [h]	No	Pumbed
Scienceseeker	scienceseeker	scienceseeker:blog_posts	No	ScienceSeeker
Scopus citations	scopus	plosalm:scopus [i]	No	Scopus
Wikipedia	wikipedia	wikipedia:mentions	No	Wikipedia
Delicious	No	delicious:bookmarks	cited by delicious count	Delicious
Facebook	facebook_shares	facebook:shares [j]	cited by fbwalls count	Facebook clicks, comments, likes
Mendeley readers	mendeley shares	mendeley readers [k]	mendeley readers	Mendeley readers, groups
Twitter	twitter	topsy:tweets [l]	cited by tweeters count	Topsy tweets

Note:
[a] These are the exact names for each data source in the PLos ALM API.
[b] You can not request a specific source from the ImpactStory API, so these are the names of the fields in the returned JSON from a call.
[c] You can not request a specific source from the Altmetric API, so these are the names of the fields in the returned JSON from a call.
[d] Some of these names are the exact names returned in an API call; others are not.
e Collected from the PLoS ALM API.
f PLoS ALM also provides xml_views.
g Collected from the PLoS ALM API. Other PMC data fields collected from PLoS ALM (pmc_abstract, pmc_supp-data, pmc_figure, pmc_unique-ip) and from PubMed (suppdata_views, figure_views, unique_ip_views, pdf_downloads, abstract_views, fulltext_views).
h Should be equivalent to plosalm:pubmed_central. ImpactStory also collects pubmed:pmc_citations_reviews f1000 pmc_citations_editorials.
i Collected from the PLoS ALM API. Scopus citations also collected from Scopus itself, in the field scopus:citations.
j ImpactStory also collects Facebook clicks, comments, and likes.
k ImpactStory also collects Mendeley readers by discipline, number of groups that have added the article, percent of readers by country, and percent of readers by career_stage.
l ImpactStory also collects the number of influential_tweets from Topsy.

Article-level metrics data provenance

Article-level metrics data comes from somewhere—tweets from Twitter, citations from Web of Science or Scopus, bookmarks from Citeulike, etc. Provenance is concerned with the origin of

an object, the ability to trace where an object comes from in case there is any need to check or validate data.

Why should we care about provenance in article-level metrics? In any research field, the verifiability of research results should be a priority, and verification requires the underlying data. Second, in general, article-level metrics are based on completely digital data. This means that all use of, research on, hiring decisions based on, and conclusions drawn from article-level metrics data should theoretically be traceable back to the original production of that data. This is somewhat unusual; most research fields are based on data collected at some point that cannot be traced, but this should be possible in article-level metrics. A specific example will demonstrate the power of data provenance in article-level metrics. Imagine if a research paper makes controversial claims using article-level metrics data on a set of objects (e.g., scholarly papers). An independent researcher could theoretically drill down into the data collected for that paper, gain further insight, and potentially dispute or add to the latter-mentioned paper.

As previously discussed, data for the same article-level metrics resource could be calculated in different ways and collected at different times for the same object. The providers already provide the date the metrics were updated. However, there is little information available, via their APIs at least, regarding how data were collected and what, if any, calculations were done on the data. The article-level metrics community overall would benefit from transparency in how data are collected.

There are two types of ways to track provenance: via URLs and identifiers. ImpactStory provides a field named *provenance_url* with each metric data source. For example, for a recent paper by Piwowar et al. with DOI 10.1371/journal.pone.0000308,[23] a GET call to the ImpactStory API returns many metrics, one of which is 10 bookmarks on Delicious. Importantly, they also return the field *provenance_url* (in this case http://www.delicious.com/url/9df9c6e819aa21a0e81ff8c6f4a52029), which takes you directly to the human readable page on Delicious from where the data was collected. This is important for researchers to replicate and verify any reported results. A nice feature of digital data such as article-level metrics is the ability to trace back final article-level metrics from providers such as ImpactStory to their original source.

The PLOS ALM API provides something less obvious with respect to provenance, a field called *events_url*, which for the same Piwowar et al. paper returns 82 bookmarks on Citeulike, and the human readable link to where the data was collected (http://www.citeulike.org/doi/10.1371/journal.pone.0000308).

Plum Analytics does something interesting with respect to provenance. In addition to the canonical URL, they collect alias URLs for each object for which they collect metrics. For example, for the DOI 10.1371/journal.pone.0018657,[24] they collect many URLs that point to that paper. This makes sense as a digital product is inevitably going to end up living at more than one URL (the internet is a giant copying machine after all), so collecting URL aliases is a good step forward for article-level metrics. ImpactStory and Altmetric (except for Mendeley) do this as well.

An important issue with respect to provenance is that data sources sometimes do not give URLs. For example, CrossRef and Facebook don't provide a URL associated with a metric on an object. Therefore, there is no way to go to a URL and verify the data that was given to you by the article-level metrics provider.

All four providers collect multiple identifiers, including DOI, PubMed Identifier (PMID), PubMed Central ID (PMCID), and Mendeley UUID. These identifiers are not URLs but can be used to track down an object of interest in the respective database/service where the identifier was created (e.g., a DOI can be used to search for the object using CrossRef here http://crossref.org/, or appending the DOI to http://dx.doi.org/).

What is ideal with respect to data provenance? Is the link to where the original data was collected enough? Probably so, if no calculations were done on the original data before reaching users. However, some of the providers do give numbers which have been calculated. For example, ImpactStory puts some metrics into context by calculating a percentage relative to a reference set. Ideally, how this is done should be very clear and accessible.

Putting article-level metrics in context

Raw article-level metrics data can be, for example, the number of tweets or the number of html views on a publishers website. What do these numbers mean? How does one paper or dataset compare to others? ImpactStory gives context to their scores by classifying scores along two dimensions: audience (scholars or public) and type of engagement (view, discuss, save, cite, recommend). Users can then determine whether a product (paper, dataset, etc.) was highly viewed, discussed, saved, cited, or recommended, and whether by scientists or by the public. This abstracts away many details; however, users can drill down to the underlying data via their API and web interface.

Altmetric uses a different approach. They provide context for only one metric, the altmetric score. This is a single aggregate metric, the calculation of which is not explained. They do provide context for the altmetric score, including how it compares to: a) all articles in the same journal, b) all articles in the same journal published within three weeks of the article, c) all articles in the Altmetric database, and d) all articles in the Altmetric database published within three weeks of the article. Altmetric gives detailed context for some article-level metrics, including Facebook, Twitter, and blogs.[25]

Plum Analytics does not combine article-level metrics into a single score as does Altmetric, but does categorize similar types of article-level metrics into captures, citations, social media, mentions, and usage (though you can dive into the individual article-level metrics).[26]

One of the advantages of article-level metrics is the fact that they measure many different things, important to different stakeholders (public, scholars, funders). Thus, combining article-level metrics into a single score defeats one of the advantages of article-level metrics over the traditional journal impact factor, a single metric summarizing data on citations. The single Altmetric score is at first appealing given its apparent simplicity. However, if article-level metrics are to avoid the pitfalls of the Journal Impact Factor,[4] we should strive for meaningful article-level metrics important to different stakeholders, that retain their context (e.g., tweets vs. citations).

A specific example highlights the importance of context. A recent paper of much interest titled *Glass shape influences consumption rate for alcoholic beverages*[27] has, at the date of this writing, an Altmetric score of 316; this score is compared relative to the same journal (PLOS One) and all journals at different points in time. Other article-level metrics are reported but are not given any context. ImpactStory reports no single score, gives raw article-level metrics data,

and gives context. For example, ImpactStory reports that there are 149 tweets that mentioned the paper and this number of tweets puts the paper 97th-100th percentile of all Web of Science indexed articles that year (2012). This context for tweets about an article is more informative than knowing that the paper has an Altmetric score of 316 —data consumers should know the context of the audience the tweets represent. The number of tweets relative to a reference set gives a bit of information on the impact of the paper relative to others. Of course not all journals are indexed by Web of Science and the important reference set for one person (e.g., papers in journals in their specific field) may be different from another person's (e.g., papers for colleagues at their university or department). PLOS recently started reporting "Relative Metrics" in the html versions of their articles, where one can compare article usage (cumulative views) to references sets of articles in different fields.[28].

There is still work to do with respect to context. Future work should consider further dimensions of context. For example, perhaps users should be able to decide how to put their metrics into context. Instead of getting raw values and values relative to a pre-chosen reference set, users could choose what reference they want to use for their specific purpose. In addition, but much harder to achieve, is sentiment or the meaning of the mention. That is, was a tweet or citation about a paper mentioned in a negative or positive light?

Historical context

Researchers asking questions about article-level metrics could ask more questions specifically dealing with time if historical article-level metrics data were available. PLOS provides historical article-level metrics data on some of their metrics (except in case of licensed resources, e.g., Web of Science and Scopus), while Altmetric provides publicly available historical data on their Altmetric score and historical data on other metrics to commercial users, and ImpactStory and Plum Analytics do not provide historical data. The data returned, for example, for number of tweets for an article from ImpactStory, Altmetric, or Plum Analytics is a cumulative sum of the tweets mentioning that article. What were the number of tweets mentioning the article one month ago, six months ago, one year ago? It is a great feature of PLOS ALM that you can get historical article-level metrics data. In fact, PLOS wants this data themselves for things like pattern detection and anti-gaming, so providing the data to users is probably not much additional work. However, these historical data are only available for PLOS articles.

The article-level metrics community would benefit greatly from storing and making available historical article-level metrics data. However, as more products are tracked, historical data will become expensive to store, so perhaps won't be emphasized by article-level metrics providers. In addition, a technical barrier comes in to play in that pushing a lot of data via an API call can get very time consuming and resource intensive.

Technical barriers to use

Some article-level metrics users may only require basic uses of article-level metrics, like including article-level metrics on their CVs to show the various impacts of their research.[13] Some users may want to go deeper and perhaps collect article-level metrics at finer time scales, or with more detailed data, than are given by article-level metrics aggregate providers. What are the barriers to getting more detailed article-level metrics data?

Diving deeper into article-level metrics means considering whether one can access data, whether the data source is machine readable, and how easy the data is to retrieve and manipulate once retrieved.

- **Data access** – Many article-level metrics sources are accessible as the data providers have open, or at least partly open, APIs (e.g., PLOS). Other data sources are problematic. For example, you can only get tweets from Twitter for the past 30 days, after which point you have to pay for a service that caches historical Twitter data (e.g., Topsy). Others are totally inaccessible (e.g., Google Scholar citations).

- **Machine readable** – Ideally, article-level metrics are provided through an API. However, some metrics of interest may only be in PDFs, spreadsheets, or html, which cannot be easily machine-consumed and re-used or mashed up. For these metrics, the user should seek out aggregators such as those discussed in this paper to do the heavy lifting. Alternatively, technically savvy researchers could write their own code, or leverage tools such as ScraperWiki.[29]

- **Ease of use** – Fortunately, many libraries or extensions exist for a number of programming languages relevant to scholars (Python, R), which deal with interacting with article-level metrics data (e.g., Figshare API libraries, Twitter API libraries; see Table 1). These libraries take care of the data collection and transform data to user friendly objects, allowing users to do the real science work of analysis and inference.

Conclusion

Article-level metrics measure the impact of scholarly articles and other products (e.g., datasets, presentations). These measures of scholarly impact are quickly gaining ground as evidenced by the four companies aggregating and providing article-level metrics (see Table 1). In any field growing pains are inevitable; article-level metrics as a field is quite young and, therefore, has some issues to work out. As shown in this paper, article-level metrics users should consider a variety of issues when using article-level metrics data, particularly consistency, provenance, and context. Article-level metrics providers collect data at different times and from different sources; combining data across providers should be done with care. Article-level metrics is special in the sense that all data is digital. Thus, there is no reason we shouldn't be able to track all article-level metrics data to their sources. This will not only provide additional insight to scholarly impact, but provide a way to verify results and conclusions made regarding article-level metrics.

As article-level metrics grow in use and popularity, researchers will ask more questions about the data. In addition, it is hard to predict what people will want to do with article-level metrics data in the future. Since we are in the early stages of the field of article-level metrics, we have the chance to steer the article-level metrics ship in the right direction. The points covered in this paper provide fodder for article-level metrics providers and users to consider.

================

Scott Chamberlain (scott@ropensci.org) is a Postdoctoral Fellow in the Biology Department at Simon Fraser University (Canada).

References

[1]   Garfield, Eugene.. "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas." *Science,* July 15, 1955, 122 (3159): 108–111. doi: dx.doi.org/10.1126/science.122.3159.108.
http://garfield.library.upenn.edu/papers/science1955.pdf

[2]   Garfield, Eugene. "The history and meaning of the journal impact factor." *JAMA: the journal of the American Medical Association*, January 4, 2006, 295 (1): 90–93.
http://garfield.library.upenn.edu/papers/jamajif2006.pdf

[3]   Monastersky, Richard. "The Number That's Devouring Science." *The Chronicle of Higher Education*, October 14, 2005. http://chronicle.com/article/The-Number-That-s-Devouring/26481

[4]   The PLOS Medicine Editors. "The Impact Factor Game." *PLOS Med,* June 6, 2006, 3 (6) (June): e291. doi: http://dx.doi.org/10.1371/journal.pmed.0030291.

[5]   Rossner, Mike, Heather Van Epps, and Emma Hill. 2007. "Show me the data." *The Journal of Cell Biology,* December 17, 2007, 179 (6): 1091–1092.
doi: http://dx.doi.org/10.1083/jcb.200711140

[6]   Alberts, Bruce. "Impact Factor Distortions." *Science,* May 17, 2013, 340 (6134): 787.
doi: http://dx.doi.org/10.1126/science.1240319.

[7]   Priem, J., D. Taraborelli, P. Groth, and C. Neylon. 2012. *Altmetrics: A Manifesto*. V 1.01, September 28, 2011. http://altmetrics.org/manifesto/

[8]   Yan, Koon-Kiu, and Mark Gerstein. "The Spread of Scientific Information: Insights from the Web Usage Statistics in PLoS Article-Level Metrics." *PloS ONE, May 16, 2011,* 6 (5): e19917. doi: http://dx.doi.org/10.1371/journal.pone.0019917

[9]   Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, and Ryan Chute. "A Principal Component Analysis of 39 Scientific Impact Measures." *PloS ONE*, June 29, 2009, 4 (6): e6022. doi: http://dx.doi.org/10.1371/journal.pone.0006022

[10]   Darling, Emily S., David Shiffman, Isabelle M. Côté, and Joshua A. Drew. "The Role of Twitter in the Life Cycle of a Scientific Publication." *PeerJ PrePrints,* May 2013, 1: e16v1.
doi: http://dx.doi.org/10.7287/peerj.preprints.16v1

[11]   Eysenbach, Gunther. "Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter And Correlation With Traditional Metrics of Scientific Impact." *Journal of Medical Internet Research*, Oct. -Dec. 2011, 13 (4):e123.
doi: http://dx.doi.org/10.2196/jmir.2012

[12]   Thelwall, Mike, Stefanie Haustein, Vincent Larivière, Cassidy R. Sugimoto. "Do Altmetrics Work? Twitter And Ten Other Social Web Services." *Plos ONE,* May 28, 2013, 8 (5): e64841. doi: http://dx.doi.org/10.1371/journal.pone.0064841.

[13]   Piwowar, Heather, and Jason Priem. "The power of altmetrics on a CV." *Bulletin of the American Society for Information Science and Technology,* April/May 2013, 39 (4): 10–13. http://www.asis.org/Bulletin/Apr-13/AprMay13_Piwowar_Priem.html

[14]   Public Library of Science (PLOS)
www.plos.org/

[15]   ImpactStory
impactstory.org/

[16] Altmetric
      www.altmetric.com/
[17] Plum Analytics
      www.plumanalytics.com/
[18] CitedIn
      citedin.org/
[19] ScienceCard
      sciencecard.org/
[20] Personal communication with Andrea Michalek.
[21] Topsy
      topsy.com/
[22] Arslan, Betül K., Eric S. Boyd, Wendy W. Dolci, K. Estelle Dodson, Marco S. Boldt, and
      Carl B. Pilcher. "Workshops Without Walls: Broadening Access to Science Around The
      World." *PloS Biology,* August 2, 2011, 9 (8): e1001118.
      doi: http://dx.doi.org/10.1371/journal.pbio.1001118
[23] Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. 2007. "Sharing Detailed
      Research Data is Associated with Increased Citation Rate." *PLoS One,* March 21, 2007, 2
      (3): e308. doi: http://dx.doi.org/10.1371/journal.pone.0000308
[24] Piwowar, Heather A. "Who Shares? Who Doesn't? Factors Associated with Openly
      Archiving Raw Research Data." *PLoS ONE*, July 13, 2011, 6(7): e18657.
      doi: http://dx.doi.org/10.1371/journal.pone.0018657
[25] See for example: http://altmetric.com/details.php?citation_id=1270911
[26] Overview: Plum Metrics, Current List of Metrics [webpage]. http://www.Plum
      Analytics.com/metrics.html
[27] Attwood, Angela S., Nicholas E. Scott-Samuel, George Stothart, and Marcus R. Munafò.
      "Glass Shape Influences Consumption Rate For Alcoholic Beverages." *PloS One,* August
      7, 2012, 7 (8): e43007. doi: http://dx.doi.org/10.1371/journal.pone.0043007
[28] Allen, Liz. "Providing context to Article-Level Metrics." PLOS Blogs, May 7, 2013.
      http://blogs.plos.org/plos/2013/05/providing-context-to-article-level-metrics/
[29] ScraperWiki
      https://scraperwiki.com/