ISQ

TOPIC

ALTMETRICS

NISO
How the information world
CONNECTS

IP

Jennifer
Lin

Martin
Fenner

Altmetrics in Evolution:
Defining and Redefining the Ontology of Article–Level Metrics

JENNIFER LIN AND MARTIN FENNER

The Public Library of Science (PLOS) first began to offer Article-Level Metrics (ALM)—including online usage, citations, and altmetrics—in 2009 to provide the research community with a view into the reach of our papers. The term altmetrics didn't exist yet, but the social bookmarks, science blog posts, and user comments collected at the time are all part of the discussion of scholarly articles on the social web or what we now call altmetrics. With ALM the emphasis is on research output in the form of an article instead of the journal that aggregates sets of articles.

Since 2009, the data sources in the PLOS suite, as well as the introduction of third party services that have joined us in aggregating altmetrics and ALMs, have experienced an upsurge. Today we have more ways to capture engagement with research outputs and more providers operating in this space than ever before. As a result, the existing landscape of ALMs and altmetrics is increasingly difficult to manage, understand, and navigate. It has become obvious that the different metrics we group together under the broader term altmetrics are indeed representing very different things. A tweet or Facebook "like" of a paper has different meaning from a user adding a paper to his/her Mendeley library or from a blog post discussing a paper. This article is borne out of such a dilemma and offers an approach aimed at

alleviating what William James called the "blooming, buzzing confusion" as the scholarly community continues to develop the new technologies into a mature and formal part of the research assessment infrastructure.

Indeed, altmetrics hinges on the very prevalence of its own diversity. Its raison d'être is to provide a more expansive view of a research artifact's impact. Put differently, the circumstance that James has imparted is in fact the very condition of the existence of altmetrics (and core to their value). We are, in fact, the very baby that James describes in the quote who, in newly experiencing the world, is assailed by a whole host of sensations from discrete objects without organizational or conceptual association. To chart a future course for altmetrics, we need
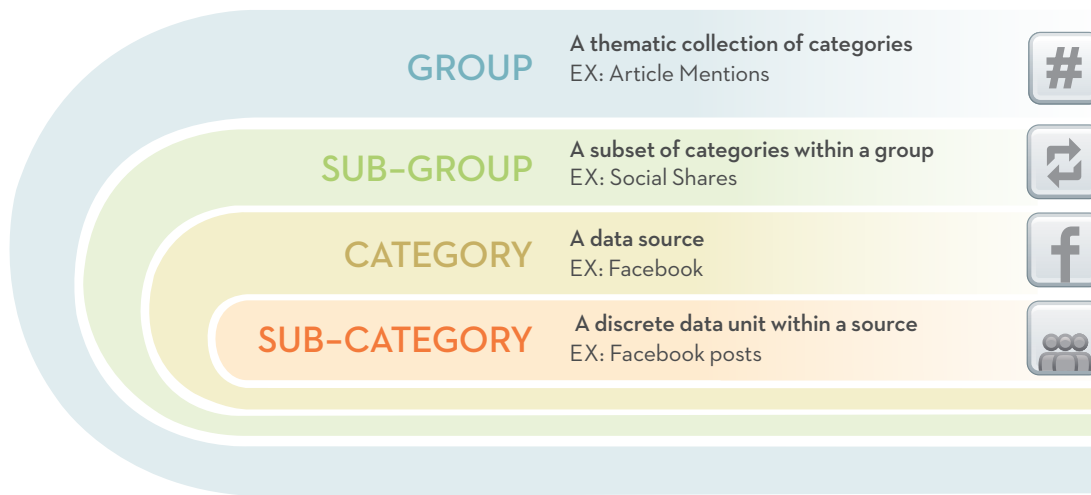
Figure 1: ALM Taxonomy

to organize the myriad metrics and make them trustworthy for all possible uses in research assessment. One important aspect of this is our ability to establish thoughtful and meaningful ways of grouping similar altmetrics together and distinguishing them from other altmetrics with different meaning. These groupings have to be used across the research ecosystem (by researchers, funders, research institutions, and publishers alike) and need to be sufficiently accommodating to endure the evolution of the assessment technology over time. We need classifications that function as infrastructure, governing how we understand and use the metrics.

We have endeavored to address this need for the purposes of PLOS's own use of its ALMs and, more broadly, to ensure that this new paradigm of assessing research takes root. The original groupings established in 2009 were no longer supporting the breadth of metrics now offered and were not in synch with those from other altmetrics providers who have since emerged. We embarked on a process of reconstructing specifications for groupings, which, broadly speaking, were made up of three overall components: evaluation, classification, and implementation. To start, we established a controlled vocabulary to reference the entities and each of their variations, as well as teasing out guiding principles for classification. Next, we evaluated the natural affinities between metrics for common groupings to arise in a manner native to the data sources. From this set of classifications we then established a framework concerning

their use throughout the PLOS journals and implemented the applications of ALM data.

Evaluation

We began with a handful of metrics at the start of the program, which were made up mostly of citations, online usage, and social bookmarking data. Over time, we have expanded the number and type of ALMs—e.g., by adding social media metrics from Twitter and Facebook—and have identified more areas to continue this escalation. But we felt that we needed to take stock and formally characterize the metrics by type and subtype at a certain point. We initiated an effort to develop a standard taxonomy of terms to take into account the different dimensions of common affinities possible amongst the diverse data, as shown in Figure 1.

The taxonomic levels primarily serve a formal mechanism of delineating the different types of metrics. The generic tokens, "metric" and "ALM," can refer to any and all of them. Any confusion arising between them only further complicates our attempts to determine suitable classifications. So we established a working taxonomy not only to establish a more precise vocabulary, but also to identify fundamental differences between the minimum component (a sub-category) and all the larger entities that include it.

In addition to the distinctions made within this taxonomy of terms, ALMs can also be characterized as

primary or secondary metrics. The former set includes the raw counts of activity captured by each source, while the latter is comprised of descriptive statistics that give context to the primary metrics (e.g., article view to PDF download ratio and average usage of similar papers). Moreover, this burgeoning set of metrics can be further distinguished based on the level of the entity measured (research paper or its component sections, e.g., figures, individual sections, etc.), type of artifact measured (article, presentation, dataset, etc.), and entity of interest (article, researcher, institution, funder, etc.). We bracketed out the latter set of distinctions to start and ascertained the broader characteristics of the very basic model.

We then established a set of general principles based on the nature of the data sources and activity captured. They emerged out of the taxonomy and the relationships between groups outlined by it.

① The grouping should be comprehensive such that each discrete metric can be placed in one and only one group.

② The grouping should ideally be structured at a level that accommodates new ALMs in the future (and flexibly named as such).

③ The grouping should ideally cluster ALMs together that share the following traits:
» Temporality
» Correlation of activity (count) to other ALMs
» Correlation of native format (e.g., event with date, title, author) to other ALMs

④ Not all the metrics for a grouping will necessarily be represented together in every aggregate. While aggregates (roll-ups) will usually align with groupings, they do not have to include all sources within each group.

REDIRECTION

Our approach ultimately consisted of a single determinant: the purpose and nature of measurement. We moved from an emphasis on the data source itself to the underlying activity captured by the data source.

These principles not only guided the classifications process but also served to "ground" an effort that involved distilling constants in the midst of continual change from the still-evolving, ever-proliferating data sets and sources. They were also incorporated into the methodology so as to avoid bias in the determination.

Classification

We began the process by setting aside the existing groups of article usage, citations, social networks, blogs and media coverage, and PLOS readers. The categories, once responsive and informative, had become rigid and mute structures that no longer reflected deep commonalities. The internal tensions between metrics within classifications had increased as new metrics were introduced. These then amplified the overall conceptual vulnerabilities of the classification system.

Thus, the slate was cleaned and we began anew. Our approach ultimately consisted of a single determinant: the purpose and nature of measurement. We moved from an emphasis on the data source itself to the underlying activity captured by the data source. The original groups were generalizations of the counts included in a group, so that social media sources were lumped together, for example. But we returned to the basic premise of ALMs and what they offer: a view into the impact and reach of an article by measuring the degree of engagement with it. With this cornerstone, we shifted to the type of article activity as the basis of establishing classifications.

Online usage is the first step of user engagement as it captures the initial (direct) encounter with the paper. PLOS tracks HTML pageviews of fulltext articles (there are no abstract pages) as well as PDF (and XML) downloads.
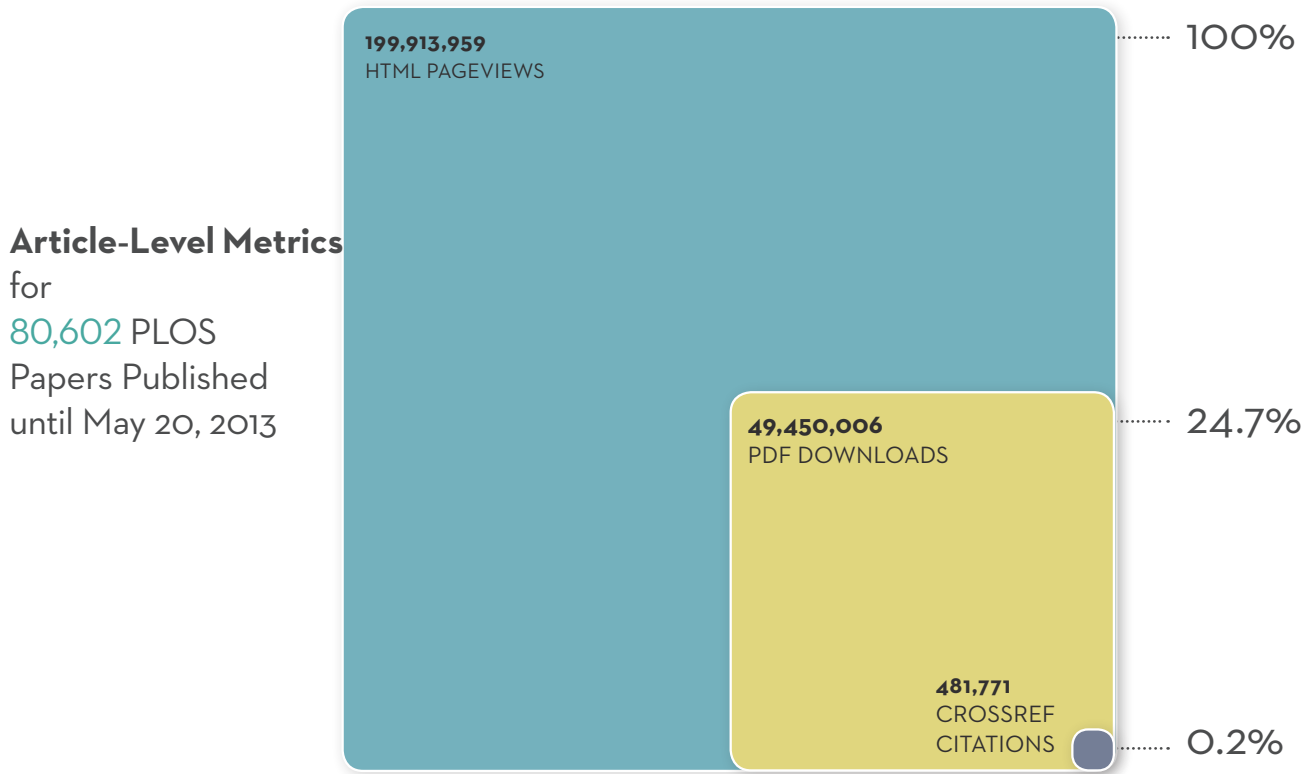
We combine the activity captured on our site with that of PubMed Central, a disciplinary repository, where fulltext copies of all PLOS articles are made freely available. On the other end of the user engagement spectrum are citations in the scholarly literature, which are tracked via the citation indices from CrossRef, Web of Science, Scopus, and PubMed Central.

Citations might be the most important measure of impact, but they only represent a small fraction of the user engagement with a paper, as shown in Figure 2. Only about one in 70 users who download a PDF of the paper will cite it. But many more will engage with it in other ways, and some of this activity can be captured with altmetrics.

When we examined the types of engagement captured by the data sources and grouped them together, we noticed a natural accession of increasing interest in and level of

Figure 2: PLOS ALM Comparison of Usage, Downloads, and Citations

Article-Level Metrics
for
80,602 PLOS
Papers Published
until May 20, 2013

199,913,959
HTML PAGEVIEWS

49,450,006
PDF DOWNLOADS

481,771
CROSSREF
CITATIONS

100%

24.7%

0.2%

RESEARCH ARTICLE

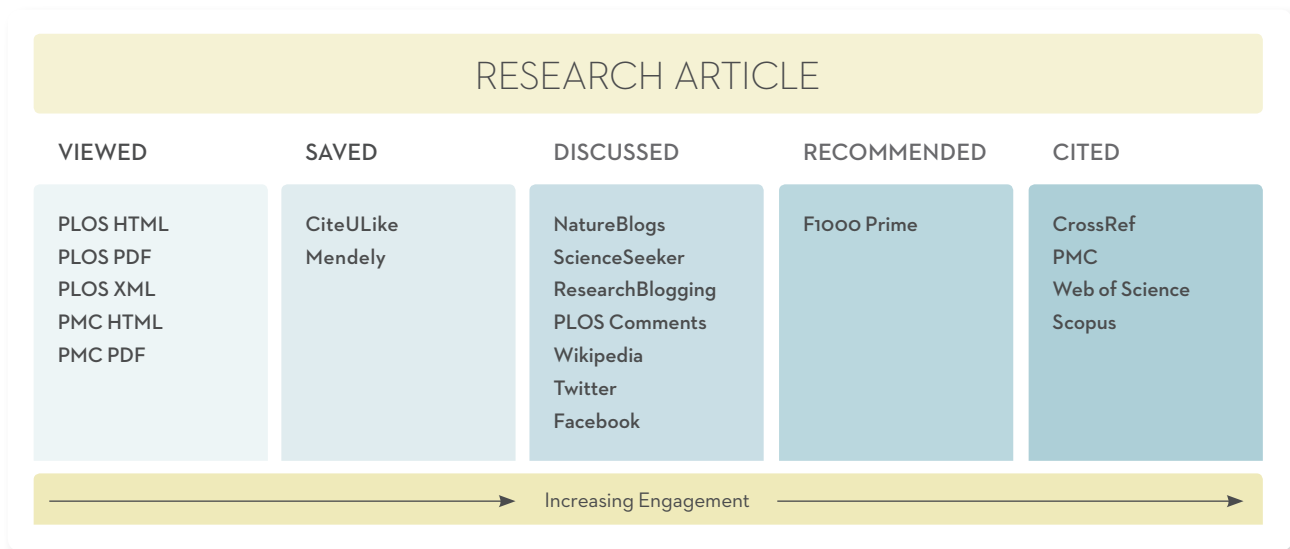| VIEWED	SAVED	DISCUSSED	RECOMMENDED	CITED
PLOS HTML PLOS PDF PLOS XML PMC HTML PMC PDF	CiteULike Mendely	NatureBlogs ScienceSeeker ResearchBlogging PLOS Comments Wikipedia Twitter Facebook	F1000 Prime	CrossRef PMC Web of Science Scopus

Increasing Engagement

Figure 3: PLOS ALM Classifications

engagement with the research articles. These fall into the following groups:

① **VIEWED**: Activity of users accessing the article online.

② **SAVED**: Activity of saving articles in online bibliography managers, which helps researchers organize papers for themselves as well as share them with others.

③ **DISCUSSED**: Discussions of the research described in an article (ranging from a short comment shared on Twitter to more in-depth comments in a blog posting).

④ **RECOMMENDED**: Activity of a user formally endorsing the research article (via a platform such as an online recommendations channel).

⑤ **CITED**: Formal citation of an article in other scientific journals.

These groups, summarized in Figure 3, are meaningful not only in that they are coherent in themselves and between each other, but also inasmuch as they reflect shared correlations to other metrics. Priem, Piwowar, and Hemminger's study offers observations that agree with the recommended groupings. Furthermore, we aimed to establish a scalable ontology that will provide affordances for the continued introduction of future ALMs (e.g., database links, news media coverage, repository links, sub-article components, etc.).

We examined the classifications by other altmetrics aggregators to uncover natural affinities present between the four today. Plum Analytics organizes their suite of metrics with the categories of usage, captures, mentions, social media, and citations. These loosely correspond to our former set, but the new groups are more closely aligned with those from ImpactStory (see Figure 4), an ontology that largely influenced ours.

As is evident from Figures 3 and 4, the key difference between PLOS and ImpactStory classifications hinges upon a delineation, used by the latter, between scholar and public metrics. We gave serious consideration to this approach, but decided that while there is a great need to be able to better assess the "people behind the data" or, more specifically, the level of significance carried by the activity captured, these distinctions are not a tight fit. The metrics designated as public ones do form a superset of both scholars and non-scholars. Even within a source, we see shifts in the groups represented across time. While a paper may be viewed quite broadly between researchers and the public upon publication, researchers will represent more of the user base over the long run. We also see differences in scholarly vs. non-scholarly activity within a group, e.g., primarily scholarly online usage from PubMed Central vs. online usage by scholars and non-scholars at the PLOS website. We hope to develop more sophisticated technologies in the future, offering deeper insight into the demographics of the users

whose article engagements are captured by the metrics, including scholar vs. non-scholar, but also by geography, career stage, etc. Until then, we have elected not to establish a public metric that is segregated from the purview of scholarly activity.

Implementation

Once the ALM ontology was established, the classifications were propagated for use in the PLOS journals. We sought to create overall consistency and coherence for the suite of metrics. But we continually found this effort ran up against our ability to fully deploy the metrics to support research discovery and evaluation of our content. The classifications gave us rules enabling us to systematically organize the metrics in logical groups as well as make them more convenient, portable, and easy to use. But we found there was a need to either group or name them differently at times, depending on the use case at hand. This recurring dilemma was expressed as a choice between overall consistency or maximum usability.

To address this issue, we have constructed a theoretical distinction at the heart of this tension between ALMs and the application of ALM data. From the perspective of the "consumer" of the data (i.e., the researcher, librarian, funder, et al.), there should be no difference between ALMs and their applications, but rather a seamless stream of real-time data supporting the navigation of the site as well as discovery and evaluation of content across the journal platform. For example, the numbers found related to an article should

agree with the ALMs used to sort search results that pull up said article.

However, the functional implementation of ALMs in PLOS journals occasionally calls for differentiating, more broadly speaking, from ALMs and their applications. ALMs come directly from the data provider (i.e., the source) and represent the activity captured in the metric. They are directly displayed most often with their primary provenance—their respective group. Conversely, we draw from ALMs as a tool to support article search and sort, assess article engagement, and report on the most popular articles. In order to apply the data to address a wide variety of possible uses, we often need to re-present it in the context of each scenario type. Here, the data is called into dialogue with the environmental factors related to each specific use case and thereby re-appropriated so as to fulfill the express purpose of the intended use.

We take a judicious and measured approach in considering modifications to the groupings and titles dictated by the classification nomenclature. In the event it is deemed necessary to fulfill a specific application, we explicitly reference the original groupings as much as possible (i.e., retain the root word). By preserving and privileging the natural base composition of ALM data through groups, we can consistently use the metrics in a fashion true to their nature (i.e., the nature of the activity captured on the article). But we can also make the ALM data "usable" by applying them to their fullest use in their application. Here, we have greater room to manipulate the display and overall form of the data while staying true to the underlying ontology at the heart of the data ecosystem. In the act of re-appropriating the data, we may manipulate the data in a number of ways, including, but not limited to, aggregating categories to fit a specific need in a way that deviates from the base group collection (sub-group) as well as modifying the grammatical state of a group or sub-group's title.

In our implementation, sub-groups are composites that operate in each instance as an expression of the data established to perform a specific function. They are comprised of a subset of categories within a group. In the event that a subgroup is expressed as an aggregate figure, each of the constitutive subgroup elements remains commensurable to the others and springs from the same type of activity captured in the metrics. All things considered, we default to the classifications nomenclature and display any assortment of ALMs based on their member grouping.

PLOS "article signposts" illustrate the distinction discussed between ALMs and their application as well as sub-groups in action. The signposts are found at the top of every article as navigational pointers for readers to get a

	SCHOLARS	PUBLIC
RECOMMENDED	Citations by editorials, f1000	Press article
CITED	Citations, Full-text mentions	Wikipedia mentions
SAVED	CiteULike, Mendeley	Delicious
DISCUSSED	Science blogs, journal comments	Blogs, Twitter, Facebook, etc.
VIEWED	PDF Downloads	HTML Downloads

Figure 4: ImpactStory Altmetrics Classifications

quick sense of the paper's flavor. The full selection includes citations, social media shares, bookmarks, and usage. (They only display when data exists for each article). Undoubtedly, they do expose the ALM data and are composed of it. And some of the signposts—citations and usage—easily correspond on the surface to existing groups. However, the signposts fundamentally are appropriations of ALM data. By securing and protecting this distinction, we have more latitude to aggregate, label, and display them so as to satisfy the purpose of providing signposts.

The signposts retain many characteristics of the groups, but minor modifications have been applied, including the inclusion of a sub-group, grammatical adjustments in the labels, etc. The signposts for the Viewed and Saved groups can be aggregated as each count represents unique activity across sources, but the Cited one must be treated differently. The four citation indices contain overlapping sets (i.e., articles that cite the respective PLOS article). Lacking a third-party open repository that de-duplicates all citations picked up by the services, this functionality called for the selection of a single data source, which would stand in for the entire set. Moreover, the signpost for the Discussed group is comprised of metrics too diverse to roll up their counts in a meaningful way. But Tweets and Facebook activity—both capturing social media activity—are similar by nature, and thus pulled out as a single number representing a sub-group to provide an additional flavor of article impact. Overall, the signposts were fundamentally constructed in deference to the groups, but modified in order to serve their purpose.

Harmonization Across ALM and Altmetric Providers

We see great potential for the role of ALMs in the discovery and evaluation of scholarly research. We have early demonstrations of their value with the PLOS implementation, and we continue to develop the program by expanding the suite of metrics as well as their applications. In these conditions, the need for ALMs is never greater than this moment when the volume of literature and other research outputs continues to exponentially skyrocket.

We are very encouraged to see a corresponding rise in the availability of ALMs for content from other scholarly publishers. With so many implementations of ALMs and altmetrics, the "buzzing, blooming confusion" we currently experience with the information overload of research content will become one of disparate metrics if the community at-large does not standardize the treatment of ALMs. As such, we see a concurrent need to harmonize the aggregation and treatment of the data across all journals and third-party providers of ALM and altmetrics data. While there seems to be overall agreement to see citations and usage stats as

groups distinct from altmetrics, there is currently no consensus on how to group altmetrics. While, for example, ImpactStory and Plum Analytics classify altmetrics sources in similar ways as PLOS, altmetric.com provides no groupings, but instead uses a single aggregate score for all altmetrics sources. As altmetrics are still relatively new to most users, these differences across altmetrics providers can create unnecessary confusion and hinder the adoption of altmetrics as a valuable addition to other metrics for research impact assessment. We at PLOS have therefore started the discussion with other providers and aggregators of altmetrics on how to group and categorize these metrics. | IP | doi: 10.3789/isqv25no2.2013.04

JENNIFER LIN (jlin@plos.org) is Senior Product Manager at the Public Library of Science.

MARTIN FENNER (mfenner@plos.org) is Technical Lead Article-Level Metrics for the Public Library of Science and Project Manager for the ORCID DataCite Interoperability Network (ODIN).

Altmetric
www.altmetric.com/

CrossRef
www.crossref.org

James, William. *Principles of Psychology.* Cambridge, MA: Harvard University Press, 1983. ISBN: 0674706250

Mendeley
www.mendeley.com/

Plum Metrics: List of Current Metrics. Plum Analytics, 2013.
www.plumanalytics.com/metrics.html

Priem, Jason, Heather A. Piwowar, and Bradley M. Hemminger. *Altmetrics in the Wild: Using social media to explore scholarly impact.* arXiv.org, March 20, 2012.
arxiv.org/abs/1203.4745

Public Library of Science (PLOS)
www.plos.org

PubMed Central
www.ncbi.nlm.nih.gov/pmc/

Scopus
www.info.sciverse.com/scopus/

Web of Science
thomsonreuters.com/web-of-science/

RELEVANT
LINKS