

Comparing digital apples to digital apples: Background on NISO's effort to build an infrastructure for new forms of scholarly assessment

Todd A. Carpenter

*National Information Standards Organization, 3600 Clipper Mill Road, Suite 302, Baltimore,
MD 21211-1948, USA
Tel.: +1 301 654 2512; Fax: +1 410 685 5278; E-mail: tcarpenter@niso.org*

Abstract. Citation metrics and the journal impact factor have been the traditional forms for decades of assessing scholarly impact. These metrics are being challenged in today's digital content environment as no longer being the best methods for research assessment due to issues such as currency and granularity. New alternative metrics – altmetrics – have sprung up, but they are not without their own issues, in particular lack of standardization. The National Information Standards Organization (NISO) launched a project in 2013 to explore, identify, and advance standards and best practices for alternative metrics. A draft white paper summarizing recommendations of the first phase has been issued and will be followed by the second phase of developing one or more of the proposed standards and recommended practices.

Keywords: Altmetrics, information standards, scientific impact, NISO, data citation, impact factor, research outputs, data discovery, research evaluation, data quality and gaming

As methods of information exchange have rapidly evolved over the past three decades, the pace of scholarship and research has accelerated. Undertaking research and sharing it with colleagues around the world is now happening electronically in a fraction of the time it used to take in a strictly print environment. Meanwhile, methods for assessment have not advanced at the same pace. The transition we have seen toward digital content distribution is also opening up a range of new opportunities for assessment of the impact of content. In order to build an infrastructure of assessment based upon the new data collection methodologies provided by digital content, we must begin with a common understanding of what should be counted, how those things will be counted, and how those resulting data should be aggregated. In short, we need a standards infrastructure for new digital assessment.

One core element of scholarly communication is the citation, whereby the author of current research acknowledges the foundational elements from past research and assumptions are recognized and referenced. In this way, researchers build on the outputs of others and create a network of interlinked results reinforcing the value of prior art and accomplishments. Everyone involved with scholarly efforts recognizes citations for their value both as a tool to support and track scientific discovery, as well as a form of recognition of the work done by others.

In the summer of 1955, Eugene Garfield, published a groundbreaking paper on citation indexing, "*Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas*" [4] in *Science* magazine. In this paper, Garfield described his vision for an information tool based on citation

studies that would allow researchers to expedite their research efforts, evaluate their work's impact, identify scientific trends, and trace the progression of scientific lines of discovery. With this paper, Garfield launched the field that would eventually come to be known as "bibliometrics" – the term coined by Alan Pritchard in 1969 [7]. Three years after this paper was published, Garfield launched the Institute for Scientific Information (ISI) with a \$500 loan to develop and calculate the Impact Factor, as it came to be known. He hired his first employee and began to build an organization that would come to employ more than 500 people when the Thomson Corporation (now Thomson Reuters) acquired ISI in 1992.

The Impact Factor and its compilation, the Journal Citation Reports [1], have become over the ensuing five-plus decades, "the" metric for assessing journal quality. Countless journals have succeeded or failed based on this metric. Many publishers in the STM community regularly tout their journal's Impact Factor in advertising and other marketing. In some developing countries, authors are awarded bonuses if they've been published in a highly ranked title [9]. For all its importance and value, the Impact Factor is an imperfect measure and the community has been arguing about its imperfections for years. Even Garfield acknowledged some of these criticisms [5], which include the time delay of citation data, the inability to compare different domains, the lack of granularity, and the figure's overuse and misapplication [8,11]. If there was one metric that the scholarly community would like to find an alternate to, it probably is the impact factor. However, for all its faults, it should be acknowledged that the Impact Factor is an ingenious and valuable metric; one that has really stood the test of time.

As noted, one of the key challenges for citation-based metrics is that they are lagging indicators of scientific inquiry. While this makes the metric generally stable and definitive based on years of consideration, it is also a hindrance in a world of near instantaneous communication. It may take years for citations to be gathered around a scholarly artifact in such a way as to make it apparent how impactful piece of scholarship has been. Most researchers, however, are interested in the latest advance or the most recent discoveries. Assessing contemporary research is an important goal of active researchers, administrators, and funders. Digital distribution and social engagements with content both provide rich sources of data upon which to draw nearer-term assessment conclusions. In addition, these newer forms of assessment also provide novel ways to assess the impact, be they temporally, quantitatively, or even qualitatively.

Perhaps with these applications in mind, the notion of finding "alternative metrics" to the Impact Factor and traditional citation measures isn't nearly as appropriate as is finding some "accompanying" or "adjunct" metrics that will add nuance to our understanding and assessment activities. While there will remain a place for traditional citation metrics, in other cases newer forms of assessment might be more valuable. Which metric is most appropriate for which situation, and which correlates most closely with eventual long-term impact are still things requiring further study.

The scope of new forms of assessment is quite broad and encompasses a tremendous range of potential metrics of impact against which scholarship can be gauged. These metrics range from determining measures of use, such as download counts, to actions to save a document in a citation manager or other research tool, to engagements with content, such as sharing it with a colleague. There are other references of content beyond its citation in traditional journal literature, such as in mass media, in legislative action, or social media – or even in books, which have been woefully neglected by citation metrics. Use can also be addressed well beyond citation, such as in adoption or application of a described approach. Other forms of less obvious use, such as network behavioral tracking or in discovery, might also have value. In addition, advances are being made and tested regarding sentiment analysis – algorithmically determining the overall tenor of a commentary – that go beyond simply the value of someone referencing something in the literature.

New metrics also need to be evaluated in the context of granularity and aggregation. Traditional metrics, such as the impact factor or COUNTER-based usage statistics [3], measure at a journal level. While this may be valuable in assessing the relative value of a publication, it is less valuable in determining the impact of a particular article, a particular researcher's portfolio, or a department's productivity. Traditional metrics were well suited to measure activity at the level of aggregation at the time, at the publication level, but they are less effective for disambiguating impact at a more granular level. In a digital environment, however, data collection can easily take place at the individual article or item level or at an even more granular level, such as figures or tables, if objects are uniquely identified, e.g. with a DOI. Other types of identifiers, such as the International Standards Name Identifier (ISNI) for institutions, ORCID's for researchers, or FundRef identifiers for funding bodies, will allow aggregation of a totally different kind than the journal package.

Additionally, many scholars are creating scholarly outputs that do not fit easily into traditional journal publication or citation models. For example many researchers are creating software, datasets, patents, conference presentations or multimedia. These non-traditional outputs have not been recognized in scholarly research evaluation because they exist outside of traditional journal publishing and, since they lack a tradition of citation, they present significant challenges to traditional metrics and have yet to develop their own methodologies for capturing use or impact.

Over the past two years, several suppliers of metrics based on the concept of alternative metrics – called altmetrics for short – have developed services around many non-traditional assessment approaches. These providers have seen significant interest and support from the community. However, several small studies [2,12] have shown that there is significant inconsistency about how such data is collected and reported, due to the lack of any standardization in this area. While this is a typical growing pain of any new methodology, it is one that must be addressed and quickly, as no community will come to rely on metrics that vary so widely from vendor to vendor or are not replicable.

The National Information Standards Organization (NISO) launched an initiative in 2013 to explore, identify, and advance potential standards and/or best practices for the community for these new forms of assessment. With support from the Alfred P. Sloan Foundation, NISO is undertaking a two-phase initiative that should take roughly three years to complete. The first phase consisted of three in-person brainstorming meetings and a round of one-on-one interviews. The three meetings were held in October 2013 in San Francisco, December 2013 in Washington, DC, and January 2014 in Philadelphia. These meetings brought together publishers, researchers, funders, and librarians to discuss potential areas of focus and concerns associated with the new assessment forms. The follow-up one-on-one interviews were targeted at under-represented groups in the three face-to-face meetings. At all of the sessions, the primary focus was on what the community might do to instill trust in new metric forms before they can be widely adopted, and what specifically NISO (alone or jointly with another organization) could do to advance a solution to those problems. The conversations were very wide ranging and brought out more than 200 potential project ideas. The nine primary themes that emerged from the discussions were:

- Definitions;
- Research outputs;
- Discovery;
- Research evaluation;
- Data quality and gaming;
- Grouping and aggregation;
- Context;

- Stakeholders' perspectives (researchers, funders, publishers, general public);
- Adoption.

A draft white paper has been issued by NISO for comments [6], which describes these nine themes in more detail and puts forward twenty-five potential project ideas culled from the hundreds that were suggested. NISO is undertaking a broad publicity effort for the draft white paper to make the results of the effort widely known and considered. Talks are planned for a number of community events throughout the summer of 2014. Following the public comment period, a survey of interested parties, as well as the application of a “thumbs up/thumbs down” voting/data gathering process, will be conducted. This community input will be used to issue a final version of the white paper and the results will be presented to NISO's Business Information Topic Committee, which has oversight for this project. They will review the results and recommend advancing one or more of the potential projects to the NISO Voting membership for final approval to launch the new projects. If approved, NISO will begin launching in the fall of 2014 the relevant Working Groups to undertake development of the proposed standards or recommended practices. It is anticipated that drafts for trial use or comment will be available in late 2015. The ultimate goal is to have final recommendations or standards published in mid-2016.

For any metric to be trusted, it must rely on consistent, verifiable data, with an understandable algorithm for its calculation. At the moment, most of the new altmetrics are being calculated on at least two of these criteria. While there remains much potential for innovation, exploration and consistency across data aggregators, enough elements are in place for standardization of these metrics to begin. NISO's community-led effort will aid in determining which aspects of new metrics are most ready for consensus-based standards or best practices. With so many potential project ideas and so much of this landscape still in flux, these initial standards can only be seen as a first step. Many more projects and avenues of research, study, testing, and further consensus will need to follow upon these first efforts. With this long-term vision of the path before us, it seems appropriate to quote the Chinese philosopher Lao Tzu, “A journey of a thousand [miles] begins with a single step” [10]. With the completion of the first phase of this initiative, a large first step has been completed. There are many more before us.

References

- [1] 2012 Journal Citation Reports[®], Science edn, Thomson Reuters, 2013.
- [2] S. Chamberlain, Consuming article-level metrics: Observations and lessons from comparing aggregator provider data, *Information Standards Quarterly* **25** (2013), 4–13.
- [3] COUNTER code of practice for e-resources, Release 4, COUNTER, April 2014, available at: http://www.projectcounter.org/code_practice.html.
- [4] E. Garfield, Citation indexes for science, *Science* **122** (1955), 108–111.
- [5] E. Garfield, Journal impact factor: a brief review, *CMAJ* **161** (1999), 979–980.
- [6] NISO Altmetrics Standards Project White Paper [draft for comments], National Information Standards Organization, Maryland, 2014, available at: http://www.niso.org/apps/group_public/document.php?document_id=13295&wg_abbrev=altmetrics.
- [7] A. Pritchard, Statistical bibliography or bibliometrics?, *Journal of Documentation* **25** (1969), 348–349.
- [8] J.J. Ramsden, Impact factors – a critique, *Journal of Biological Physics and Chemistry* **9** (2009), 139–140.
- [9] J. Shao and H. Shen, The outflow of academic papers from China: why is it happening and can it be stemmed?, *Learned Publishing* **24** (2011), 95–97.
- [10] L. Tzu, T.T. Ching and J. Legge [translator], *Sacred Books of the East*, Vol. 39, 1891, verse 64, available at: <http://www.sacred-texts.com/tao/taote.htm>.
- [11] J.K. Vanclay, Impact factor: outdated artefact or stepping-stone to journal certification?, *Scientometrics* **92** (2012), 211–238.
- [12] Z. Zahedi, M. Fenner and R. Costas, How consistent are altmetrics providers? Study of 1000 PLOS ONE publications using the PLOS ALM, Mendeley and Altmetric.com APIs, *figshare* (2014), available at: <http://dx.doi.org/10.6084/m9.figshare.1041821>.