

Metrics and Assessment

Based on summaries of two presentations from the 38th UKSG Annual Conference, Glasgow, March 2015

An important and timely plenary session at the 2015 UKSG Conference and Exhibition focused on the role of metrics in research assessment. The two excellent speakers had slightly divergent views. Todd Carpenter from the National Information Standards Organization (NISO) argued that altmetrics are not alt(ernative) any more and that downloads and other forms of digital interaction, including social media reference, reference tracking, personal library saving and secondary linking activity now provide mainstream approaches to the assessment of scholarly impact. James Wilsdon is Professor of Science & Democracy in the Science Policy Research Unit (SPRU) at the University of Sussex and is Chair of the Independent Review of the Role of Metrics in Research Assessment commissioned by the Higher Education Funding Council for England (HEFCE). The outcome of this review will inform the work of HEFCE and the other UK higher education funding bodies as they prepare for the future of the Research Excellence Framework (REF). He is more circumspect, arguing that metrics cannot and should not be used as a substitute for informed judgement. This article provides a summary of both presentations.

Todd Carpenter's session focused on the work that the National Information Standards Organization (NISO) is undertaking to look at new forms of assessment. His summary is published below.

Altmetrics aren't alt any more: altmetrics meet the mainstream

Metrics are a vital element in the exchange of information. We are now at a place where we can gather more robust information, at a much more granular level, from many more sources than ever before. We are only now beginning to test the boundaries of what we can do with these data and understand the relationship between these signals of interaction and whether they translate into eventual impact down the road.

When people think about 'alternative', many think of the 1980s 'alternative' bands, such as New Order, or The Smiths. But, actually, some of the styles from that era had throwbacks to the 1920s. So when presented with modern versions of 'alternative' styles, we may not consider them as alternative any longer. To us, they have entered the mainstream.

The same could be said for alternative metrics, or altmetrics. Those engaged in the altmetrics community have spent the past five years talking about new types of assessment metrics, getting people excited about the term, excited about what it means, and the potential of altmetrics. There have been hundreds of articles and posts about altmetrics. There have been eight conferences specifically focused on alternative metrics and there are at least four organizations focused on providing alternative-style metrics. Essentially, when the standards guy is in front of you talking about this, it is not alternative any more.

In the early days of alternative metrics, many people viewed them as simply focusing on social media references, such as Facebook 'Likes', tweets on Twitter, or blog posts. The altmetrics community faced the question of why these new metrics matter. Subsequently, there has been research published that points to alternative metrics as being good leading indicators of future citations. This research has shown that there is a modest, positive



TODD CARPENTER
Executive Director
National Information
Standards
Organization (NISO)

'Metrics are a vital element in the exchange of information'

correlation between current social media interest and future impact as measured through citation. While useful, there is a broader context for scholarship today that includes more than traditional articles published in journals. Scholarship is being distributed via a much broader range of outputs. Scholars are producing software, visualizations, patents, trademarks, video, software code, scholarly data sets, online courses and legislative testimony. Even popular media is picking up some of the science that is being produced and is distributing it. In this context, we should capture as much of that information as possible to provide a full-on view of what scholarship and what researchers are now providing and the impact of that. Journal citation is simply not the only – nor even the best – way to capture those interactions.

'there is a modest, positive correlation between current social media interest and future impact as measured through citation'

So how can we measure the impact of these different forms and types of scholarly output? In this case, we are interested in how scholarship is being received and accepted by the community. These new metrics are simply different ways of understanding the scholarly landscape. This is why the focus of the presentation was not to promote the use of altmetrics but rather to focus on the fact that these are just metrics. We should drop the term altmetrics and just talk about metrics. The reality is that our community has already been relying on non-citation-based metrics for a long time.

For example, in the year 2001, Project COUNTER was launched. At that time, the community spent a lot of time talking about article download activity and what counts as a download. There were discussions about whether a PDF and an HTML download counted as the same and how long between server requests should count as the same request. Over time, through COUNTER's promotion and adoption, article download activity became a traditional metric that is widely used by many institutions. Almost every academic publisher is using COUNTER statistics and almost every library is interested in receiving them from publishers.

But, prior to that, libraries were creating data about the use and application of their content and how people use their materials. The National Information Standards Organization's (NISO's) seventh standard, published in the 1960s, was a data dictionary about how to analyse library use and activity. Things like numbers of people walking through the door, gate checks and circulation numbers could be meaningfully compared between institutions. Each of these metrics – and there are many more – shows the community how it is performing in its role of distributing content.

If we have been gathering and applying non-citation-based metrics for decades, then what has recently changed that has attracted so much attention to altmetrics? What has changed is our ability to gather and track new forms of data – very granular data – on how things are being used in a digital landscape. We are also able to aggregate and share that information in near real time.

At this point, NISO is focused on the community's inability to compare apples and apples or compare one metric from one source against another, which really is at the core of the questions around trust. What are the elements to build trust upon as it relates to new forms of assessment? We need things like common definitions of what we are counting. We need better systems identifying the content that is distributed so that we know exactly what we are referring to. We need to better understand questions about the granularity at which we are gathering data. For example, should we count content objects at a journal level, an article level or at levels not associated with a publication, such as at a grant or department level? How do we associate content objects with their authors to derive metrics related to the scholar or the researcher? Do we have a full understanding about the timescale along which we are counting things? Finally, how do we exchange the information that we are gathering in a way that can be trusted?

'How do we associate content objects with their authors to derive metrics related to the scholar or the researcher?'

35 Each of these questions comes back to standards. How should we define what should be counted and by what method? How do we identify those things using identifiers like international standard book numbers (ISBNs), digital object identifiers (DOIs) or open researcher and contributor identifications (ORCIDiDs)? What are procedures for what to count, what not to count, and how to share and exchange that information? All of these questions return to the focus on standards.

In this context, NISO launched a project to look at new forms of assessment. To support this work, NISO received a generous grant from the Alfred P Sloan Foundation. This initiative includes work in two phases. The first phase of the project focused on brainstorming and data gathering. We organized three in-person meetings and undertook a series of one-on-one interviews. The results were summarized in a white paper NISO published in June 2014.

The three meetings were organized as un-conferences to elicit conversation. We started with a series of lightning talks. Then we had participants write on Post-It notes their ideas of what the group should discuss, what problems the community is facing, and potential solutions. We then organized those ideas and voted on them to form conversation groups.

We wanted to involve people who were not able to join us in person so we reached out to about 30 people, primarily administrators and grant funders. Out of these efforts, we generated a plethora of ideas. In the proposal, we suggested that NISO might come up with about 50 ideas. In this process the participants identified more than 200, which led us to the second phase.

In the second phase, we prioritized the ideas and addressed the top priorities. Realistically, NISO is not in a position to act on all the ideas we generated. To proceed, NISO needed the community to identify which priorities to pursue. The white paper report organized those 200-plus ideas, narrowed down into 25 potential work projects, grouped into eight themes:

- *Definitions:* What are we actually talking about when we are talking about alternative metrics?
- *Assessment:* How might these forms of assessment apply to different types of research outputs and how might institutions apply them in research assessment?
- *Discovery:* How can these data be used to improve discovery systems? For example, Amazon knows who looked at this also purchased that. This is a form of data analytics based on usage. How can the research community build things like that into its discovery systems?
- *Data quality:* How can we be assured that the data being generated are valid and conform to norms agreed by the community? Should there be audits of systems to ensure data accuracy?
- *Gaming:* Whenever there is assessment, some are drawn to cheating in some to game the systems to advance their careers. But are there ways systems can be set up to capture or limit gaming?
- *Granularity:* How we group and aggregate information is complex but critical, in order to merge this information from the different available data sources.
- *Context:* Each field is different and how do we understand the context between different data or analytics models and how they apply differently in different domains?
- *Use and adoption:* We do not want to create documents and standards stacked up on a shelf that no one uses. We want people to use these metrics and systems. How do we promote adoption and use of these new best practices?

These were the themes covered in the white paper. We then asked the community to rank the importance of the 25 different project ideas on a scale of very important to not at all. Almost 90% of the respondents thought definitions were very important or important.

36 Persistent identifiers were also critical because they underlie a lot of the data gathering and aggregation that would take place in order to provide these metrics. Again, how do we focus on these metrics as they apply to other forms of research? How we calculate them is obviously of great importance.

Now that we have finished the brainstorming and prioritization efforts, NISO is now in the process of launching three different working groups to address five issues derived out of the 200 ideas. The first working group will focus on use cases and definitions, how to define what we are talking about and how these can be applied in different communities. The second group will focus on non-traditional outputs and how we can make assessment work for things like data sets and software. This group will also look at the issue and greater use of persistent identifiers in the community. The third group will focus on the technical side of strategies for improving data quality, defining some of these, determining how to calculate some of the metrics, and where we are getting the data and how we share them.

'NISO is now in the process of launching three different working groups to address five issues derived out of the 200 ideas'

To manage the three groups, we have organized a steering committee that will co-ordinate the project, which is comprised of two co-chairs from each of the working groups: Greg Tananbaum (ScholarNext Consulting and SPARC), Mike Taylor (Elsevier), Kirsti Holmes (Northwestern University), Martin Fenner (PLOS; also a consultant to NISO on this project), Mike Showalter (EBSCO), Michael Habib (SCOPUS), together with Nettie Lagace (NISO's Associate Director) and Todd Carpenter.

About 75 people expressed interest in participating in one of these three working groups, and we need to align participants and their interests. The NISO leadership appointed members to each of the three groups and they began their work in April.

For more information about the project, visit our website and download the white paper. There was a lot more in this project than NISO or any one organization could do, so I am really pleased that a variety of other players in the community are moving forward. It is our hope that through our efforts we can advance the understanding, trust and use of these new forms of metrics.

Todd Carpenter
Executive Director
The National Information Standards Organization (NISO), USA
E-mail: tcarpenter@niso.org

ORCID ID: <http://orcid.org/0000-0002-8320-0491>

James Wilsdon, as Chair of the Independent Review of the Role of Metrics in Research Assessment, considered how robust alternative metrics are and whether the UK Higher Education Funding Councils should consider them in their management of research. His final report will be published this summer, but here, providing insight into the thinking that has informed the review, and as a teaser, is the summary of this plenary session.



JAMES WILSDON

Professor of Science & Democracy
SPRU, University of Sussex

'In metrics we trust?'

Citations, journal impact factors, H-index, even tweets and Facebook likes – there are no end of quantitative measures that can now be used to assess the quality and wider impacts of research. But how robust and reliable are such metrics, and what weight – if any – should we give them in the management of the UK's research system?

These are some of the questions that are currently being examined by an Independent Review of the Role of Metrics in Research Assessment, which I am chairing, and which includes representatives of the Royal Society, British Academy, Research Councils UK and Wellcome Trust. The review was announced by David Willetts, then Minister for Universities and Science, in April 2014, and is being supported by the Higher Education Funding Council for England (HEFCE).

Our work builds on an earlier pilot exercise in 2008–9, which tested the potential for using bibliometric indicators of research quality in the Research Excellence Framework (REF). At that time, it was concluded that citation information was insufficiently robust to be used formulaically or as a primary indicator of quality, but that there might be scope for it to enhance processes of expert review.

The current review is taking a broader look at this terrain by exploring the use of metrics across different academic disciplines and assessing their potential contribution to the development of research excellence and impact within higher education, and in processes of research assessment like the REF. It is also looking at how universities themselves use metrics, at the rise of league tables and rankings, at the relationship between metrics and issues of equality and diversity, and at the potential for 'gaming' that can arise from the use of particular indicators in the funding system.

'the relationship between metrics and issues of equality and diversity'

In Summer 2014 we issued a call for evidence and received a total of 153 responses from across the HE and research community. Of these responses, 57% expressed overall scepticism about the further introduction of metrics into research assessment, a fifth supported their increased use and a quarter were ambivalent. We have also run a series of workshops, undertaken a detailed literature review and carried out a quantitative correlation exercise to see how the results of REF 2014 might have differed had the exercise relied purely on metrics, rather than on expert peer review.

Our final report, entitled 'The Metric Tide', will be published on 9 July 2015. But ahead of that, we have recently announced emerging findings in respect of the future of the REF. Some see the greater use of metrics as a way of reducing the costs and administrative burden of the REF. Our view is that it is not currently feasible to assess the quality and impact of research outputs using quantitative indicators alone. Around the edges of the exercise, more use of quantitative data should be encouraged as a contribution to the peer-review process. But no set of numbers, however broad, is likely to be able to capture the multifaceted and nuanced judgements on the UK's research base that the REF currently provides.

'if you've been pimping and priming your H-Index in anticipation of a metrics-only REF... our review will be a disappointment'

So if you've been pimping and priming your H-Index in anticipation of a metrics-only REF, I am afraid our review will be a disappointment. Metrics cannot and should not be used as a substitute for informed judgement. But in our final report, we will say a lot more about how quantitative data can be used intelligently and appropriately to support expert assessment in the design and operation of our research system.

James Wilsdon
Professor of Science & Democracy
Science Policy Research Unit (SPRU), University of Sussex, UK
E-mail: J.Wilsdon@sussex.ac.uk

ORCID ID: <http://orcid.org/0000-0002-5395-5949>

Article copyright: © 2015 Todd Carpenter and James Wilsdon. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



To cite this article:

Carpenter, T and Wilsdon, J, Metrics and Assessment, *Insights*, 2015, 28(2), 33–38; DOI: <http://dx.doi.org/10.1629/uksg.248>.

Published by UKSG in association with Ubiquity Press on 07 July 2015