

Metrics and evaluation in publishing

Nicola Gulley¹

IOP Publishing, U.K.

Expansion and assessment

Publishing an article in a reputable journal has become intrinsically linked with research, both in terms of evaluation and communication. Indeed the core objectives remain the same today as they did in the very early days of publishing when Henry Oldenburg set up the very first journal in 1665; registration, dissemination, peer review and archival record [1]. Over the centuries, research expanded and the number of journals dramatically increased to meet this growing demand to publish. With this growth in journals, librarians and other decision makers required ways to evaluate and classify them, in terms of quality and perceived value to the community, to help prioritize between the journals. With the introduction of the Science Citation Index (SCI[®]), closely followed by the Journal Citation Reports (JCR[®]), a new way of indexing and measuring the impact of research was proposed. The development of SCI[®] and JCR[®] has been well documented in the literature (for an example see [2]).

To understand the scale in the growth, particularly from the late 1940s, we can look at the growth seen in the area of STM (science, technology and medicine) publishing. In a recent report by Ware and Mabe [3] on behalf of the STM, the International Association of Scientific, Technical and Medical Publishers, it was reported that there were approximately 28100 active scholarly peer-reviewed journals publishing around 1.8–1.9 million articles per year in 2011. They also reported that, on average, the number of articles published has grown, year on year, by approximately 3%; the number of journals by 3.5% and a corresponding growth in the number of researchers of 3%. In some research areas the growth has been much higher year on year. In physics and astronomy, for example, some years have seen growth rates as high as 13% compared with the previous year.

Journal metrics: part I

With so much choice, differentiating between journals is important for researchers in deciding where to publish, librarians when deciding which journals to buy and for publishers to benchmark their journals against the competition. Given that most researchers want to publish in a journal that not only reaches their key audience, giving them maximum visibility, but is also recognized as being prestigious by others both within and outside of their research field, a widely accepted metric

¹Email: nicola.gulley@iop.org

can be attractive. The JIF (journal impact factor) is currently used as such a metric. It was introduced as part of the JCR® as one of the tools available for categorizing and evaluating journals and is effectively a measure of the number of citations of an average article published in the journal (a full definition is given at Thompson Reuters' Web of Knowledge [4]). Although it can be a useful metric, its meaning can vary from subject area to subject area, so it is important to understand what is being compared for any comparison to be meaningful. Research communities tend to have different citation habits, and from a publisher's, and authors', perspective this can cause problems when the JIF is used as a number in isolation without any reference to average JIF in a particular research field.

For example, the median JIF for a journal in mathematics is lower than that of a journal in the area of biology. As an example, one of the journals published by IOP Publishing, *Inverse Problems*, has a JIF of 1.866 in the 2012 JCR®. Over the years it has consistently been above the category median impact factor for the two categories in which it is indexed, 'Mathematics, applied' and 'Physics, mathematical' (2012 category median impact factors 0.742 and 1.138 respectively) and is highly regarded by the researchers who publish in the journal. Compare this with the median for the category biophysics at 2.605 and you can see that if *Inverse Problems* was compared with a journal below the median in biophysics without the context of where that value sat, it could be viewed in a more negative way than it deserves.

This is a very simplistic example, as the median JIF in a particular category will depend on many factors, but the point I am highlighting is that the JIF can tell you how many times on average an article in the journal will be cited, but in isolation it cannot tell you how that journal compares with others or how it is perceived by the research community. Therefore if a researcher is told that they should only publish in journals with a JIF above 2.0, it rules out very good journals in some fields, particularly where articles can accumulate as many citations outside of the JIF calculation window (2 years) as they do within it. However, the JIF continues to be used as a key metric of the success and quality of a journal. More recently, additional metrics such as the 5-year impact factor have been introduced to try and accommodate the areas that receive citations over a longer time period. Thompson Reuters have also run a number of presentations to explain how the JIF should be used ideally, and in the example referenced, they also outline a number of other metrics that can be used in conjunction with it [5]. There are also many other examples to be found in the literature and discussed elsewhere in this book.

More recently, with the move to electronic publishing in the mid-1990s, journals have also been evaluated on usage statistics such as the number of downloads (a download refers to a full-text article being downloaded by a reader). In the early days, the biggest challenge was how to benchmark these data to enable a meaningful comparison to be made between journals and publishers. In March 2002, an initiative called COUNTER (Counting Online Usage of Networked Electronic Resources) (<http://www.projectcounter.org/index.html>) was launched to set standards for recording and reporting usage statistics in a meaningful, and credible, way. COUNTER is a not-for-profit company and for a vendor to become COUNTER compliant there is a code of practice that should be followed [6]. However, there are no rules in COUNTER that govern article-level metrics, which is important in relation to new initiatives showing article-level downloads

as a metric. The data can be compiled ‘in the spirit’ of COUNTER, but there are currently no set formats at this level, so it is difficult to compare article-level downloads across different platforms. However, there is a project that is underway (the Pirus project) to look at solutions to this.

Other factors that are taken into consideration when evaluating and comparing journals include the geographical and subject-area spread of the journal editorial board, as well as the specific subject scope of the journal. This varies from journal to journal and with the specific community they represent, but generally, an internationally respected journal will aim to have representation from all subject areas covered by the scope of the journal and a good balance of all the geographical areas where related research takes place. When evaluating a journal the geographical distribution of authors can also be a consideration.

Journal metrics: part 2

So far I have outlined what external metrics, both hard and soft, are used in assessing a journal, but there is another side to the story. With the significant growth of published articles, there have also been significant developments on the publishing side to monitor, manage and analyse the data that are accumulated as part of the infrastructure that supports publishing an article. At a publisher level, data on the geographical distribution of authors in journals are routinely collected as well as analyses of trends in journals. At IOP Publishing we run the peer review in-house for a number of journals with dedicated scientific editors. We maintain a database of a few hundred thousand referees for use across the journals using our systems. Internally, we can use these data to identify where we need more referees in particular countries or in particular subject areas. We collect data both at an article level and a journal level that is used to help develop the strategy for the journal. The editorial board members and editorial staff of the journals can review journal data including:

- The number of articles submitted to the journal over a time period
- The number of articles that do not make it through to publication
- Rejection/acceptance rates of the journal
- Processing times at different stages of the peer-review and publication process
- The number of downloads
- The number of citations
- The number of referee reports
- Geographical breakdown of authors/referees/readers
- The subject distribution

All of these data paint a detailed picture of a journal and supports the discussions around journal policies and development.

In addition to the specific development of journals, publishers can also provide some of these data at an institutional level that enables a librarian/information scientists to build a better picture of the use of research published by their institute. There are different levels of information available. For example,

Elsevier has a service which integrates external information with the data it holds on its Scopus database, called SciVal, but many other publishers also offer bespoke reports based on the data they collect. It is an area that is developing as publishers, authors and research institutes review data available in addition to the traditional metrics that can help them develop a better idea of the impact of research and the success or failure of an investment. That investment can be anything from a technical development through to the output of a research grant; it comes back to the increasing pressure to find ways of reporting and evaluating on research investment in general.

There are also broader industry initiatives that provide additional data that help in this respect. One example is CrossRef (<http://www.crossref.org/>), which is an independent membership association founded and directed by publishers. CrossRef enables the linking of references in online documents that has greatly improved the navigation between published literature (predominantly peer-reviewed journal articles). It refers to itself as 'a sort of digital switchboard' which is an excellent description of how the links are managed as CrossRef does not hold full-text content, but facilitates the linking between different publishers. Since its foundation in 2000 it has also become the official DOI (digital object identifier) link registration agency for scholarly and professional publications. DOIs act as unique and persistent identifiers of digital content and can be used for objects other than journal articles, such as datasets. This is an interesting area that is developing in terms of wider linking outside of a standard journal article as the unit. For example services such as Figshare (<http://figshare.com/faqs>) are using DOIs to link and reference data, presentations, posters, etc.

From CrossRef, there are new publishing initiatives such as FundRef (<http://www.crossref.org/fundref/>) and ORCID (<http://orcid.org/>). FundRef is aiming to standardize the way that information on funding sources for research is captured and reported by developing a standard taxonomy for funders that can be captured as part of the publication process. This can then be deposited into a central database which can be made available for analysis. ORCID is aiming to develop a registry of unique research identifiers that researchers can use to link their research activities and output and will help to disambiguate work by researchers with the same or similar names.

Future metrics

Scientific publishing is currently undergoing one of the most radical changes in its history. I started the article by stating that the core objectives for journals have not changed in close to 350 years. Registration, dissemination, peer review and a reliable archival record are still fundamental to scientific publication, but the way this is achieved is changing. We are moving from analysing the performance of the journal as a unit to the article and, in some cases, to sections of the article. With the data accumulated and facilitated as part of the publishing infrastructure and the emergence of new ways of communicating in a digital environment, there is a need to find metrics that give a wider picture of the impact of a researcher's work. Blogs are becoming more important, and there are questions about how

researchers managing vital databases get credit for the work they do when others use the data. Social media sites such as Twitter and Facebook are being used for discussions outside of the traditional channels and bookmark services such as Citeulike (<http://www.citeulike.org>) are used to index articles.

There have also been a number of new approaches to different aspects of scientific publishing. For example, *PLOS ONE* is a broad-subject journal that aims to assess whether articles submitted are technically sound before publication and facilitates comments post publication. *Atmospheric Chemistry and Physics* [7] uses a two-stage process of traditional peer review (reviewers specifically selected to review a paper) and comments collected online from an interactive public discussion. Other journals are experimenting with different formats such as video. *New Journal of Physics*, a journal published by IOP Publishing on behalf of the Institute of Physics and the Deutsche Physikalische Gesellschaft, introduced video abstracts following requests from authors for a more digital approach to explaining research in a broad-scope journal [8].

All of these different approaches require different ways of evaluation. How does a researcher get credit for a blog they write? How do you measure the impact of a video abstract? How can you measure the reach of article that has been forwarded on Twitter? Data on this are being captured, but not always presented publically. IOP Publishing, Springer, Nature Publishing Group and many other publishers are now presenting a range of data at an article level that covers citation counts, downloads, Twitter referrals, views, etc. Services such as Altmetric (<http://www.altmetric.com>) and Impactstory (<http://www.impactstory.org/about>) have been leading the way in looking at how to interpret these data. In a recent paper, Liu and Adie [9] describe how these alternative metrics can be used to complement the traditional citation analysis to build a more detailed picture of how research is used. Priem [10] goes further in his paper to discuss how this changing environment could affect scholarly publishing in the widest sense. He also addresses a number of concerns that have arisen in context of the alternative metrics being discussed, not least that they can be gamed. The metrics are also not easily benchmarked and there is currently little consistency across different websites.

A brave new world

Evaluation is useful for reviewing, measuring and improving the status quo. With more calls on resources available and limited financial constraints it becomes important to understand the outcomes of any investment. There have been a number of high-profile reviews of the processes supporting publication, from peer-review models through to wider access options such as open-access publication. Publishing and individual journals are also subject to ongoing evaluations as I have outlined above. However, there is also a need to ensure that the evaluation is clearly defined, and there are ongoing challenges with single metrics used out of context, a subject that is discussed in more detail in the bibliometrics literature than I have been able to cover in the present chapter.

Solutions have been found to new challenges as they arise, such as the COUNTER code of practice being introduced to give some standardization to the

new metric of downloads, but there is more to be done in what is a very rapidly evolving environment, not only in defining agreed benchmarks, but also in making better use of the data already being accumulated that could be used as useful new metrics in conjunction with the more traditional ones.

For researchers there is the challenge of having reliable ways of reporting the impact of their research, however that is defined. The use of new ways of communicating research can be difficult to label within the confines of the traditional scholarly communication routes, but they are complementary and need to be recognized appropriately.

There are genuine concerns about the ease with which any one metric, such as downloads at an article level, can be manipulated or within the realm of social media where a 'like' or 'dislike' can be very subjective. And, of course, there are the ever-existing concerns that any metric can be taken out of context and misinterpreted. For example, if an article is picked up by a number of people and re-tweeted or has had an associated press release that increases the web traffic and download counts to that article, does it say anything more than that the article is popular at that particular time? Does it drive citations, indicate that the article is a high-quality article or that it is a popular article? All of the above? The usefulness of an article is decided over time by the relevant research community along with how well it is cited and used, but it does not mean that the additional information that can be provided at an article level is not useful when looked at in a broader sense. Citations, download counts, social media statistics, view of a video abstract, all of these metrics potentially tell a slightly different part of the overall story and could be used together to paint a fuller picture of the research.

Article-level metrics are still relatively new and evolving quickly, their long-term usefulness still hotly debated. Challenges with the existing traditional metrics have been well documented. Of course the answer is, as always, no single metric should be used in isolation, but in addition, there is another question to be answered about how useful some of the newer metrics will be in years to come. At the time of writing, *Nature* had just published a special issue on this topic [11] covering a number of related issues. *Scholarly Kitchen* had also covered this subject on a number of occasions, most recently in a blog post by Crotty [12]. So although there are still many questions about future metrics, what is certain is that this is set to be an interesting development that will evolve quickly and will continue to be debated for some time. With the continued focus on evaluation of research it will be important that these debates involve all of the different players. Journals are subject to evaluation, but publishers are also in a position to be supplying data and the infrastructure that could be useful in supporting a different approach to measuring a researcher's contributions in the future.

Acknowledgements

I am grateful for the support of IOP Publishing in writing this article, but any opinions expressed are mine and do not represent those of my employer or anyone else.

References

1. Mabe, M. (2006) (Electronic) journal publishing. The E-Resources Management Handbook UKSG. http://www.stm-assoc.org/2006_08_23_Electronic_Journal_Publishing.pdf (Accessed 20 October 2013)
2. Garfield, E. (2006) The history and meaning of the journal impact factor. *The Journal of the American Medical Association* **295**, 90–93
3. Ware, M. and Mabe, M. (2012) *The STM report: an overview of scientific and scholarly publishing*. http://www.stm-assoc.org/2012_12_11_STM_Report_2012.pdf (Accessed 23 October 2013)
4. Thompson Reuters, Web of Knowledge. <http://wokinfo.com/essays/impact-factor/> (Accessed 21 October 2013)
5. Webster, B.M. (2009) About use and misuse of impact factor and other journal metrics. <http://wokinfo.com/media/using-the-impact-factor.ppt> (Accessed 21 October 2013)
6. COUNTER. <http://www.projectcounter.org/> (Accessed 23 October 2013)
7. Atmospheric Chemistry and Physics: aims and scopes. <http://atmospheric-chemistry-and-physics.net> (Accessed 23 October 2013)
8. New Journal of Physics: video abstracts. <http://www.njp.org/videoabstracts> (Accessed 23 October 2013)
9. Liu, J. and Adie, E. (2013) New perspectives on article-level metrics: developing ways to assess research uptake and impact online. *Insights: the UKSG Journal* **26**, 153–158
10. Priem, J. (2013) Scholarship: beyond the paper. *Nature* **495**, 437–440
11. (2013) Research evaluation: impact. *Nature*, doi:10.1038/502287a. Also available at: <http://www.nature.com/news/research-evaluation-impact-1.13949> (Accessed 23 October 2013)
12. Crotty, D. (2013) Driving altmetrics performance through marketing: a new differentiator for scholarly journals? The Scholarly Kitchen blog. <http://scholarlykitchen.sspnet.org/2013/10/07/> (Accessed 23 October 2013)