



## Aslib Proceedings

The missing link: journal usage metrics  
Ian Rowlands, David Nicholas,

### Article information:

To cite this document:

Ian Rowlands, David Nicholas, (2007) "The missing link: journal usage metrics", Aslib Proceedings, Vol. 59 Issue: 3, pp.222-228, <https://doi.org/10.1108/00012530710752025>

Permanent link to this document:

<https://doi.org/10.1108/00012530710752025>

Downloaded on: 10 May 2018, At: 02:17 (PT)

References: this document contains references to 6 other documents.

To copy this document: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)

The fulltext of this document has been downloaded 884 times since 2007\*

### Users who downloaded this article also downloaded:

(2005), "The use and users of scholarly e-journals: a review of log analysis studies", Aslib Proceedings, Vol. 57 Iss 6 pp. 554-571 <<https://doi.org/10.1108/00012530510634271>>

(2008), "What do citation counts measure? A review of studies on citing behavior", Journal of Documentation, Vol. 64 Iss 1 pp. 45-80 <<https://doi.org/10.1108/00220410810844150>>

Access to this document was granted through an Emerald subscription provided by emerald-srm:395687 []

### For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit [www.emeraldinsight.com/authors](http://www.emeraldinsight.com/authors) for more information.

### About Emerald [www.emeraldinsight.com](http://www.emeraldinsight.com)

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

\*Related content and download information correct at time of download.



AP  
59,3

222

Received 12 December 2006  
Revised 15 February 2007  
Accepted 20 February 2007

# The missing link: journal usage metrics

Ian Rowlands and David Nicholas

*Centre for Information Behaviour and the Evaluation of Research (CIBER),  
School of Library, Archive and Information Studies,  
University College London, London, UK*

## Abstract

**Purpose** – The aim of this short communication is to contribute to a growing debate about how we can measure the “quality” of journals. More specifically, the paper argues the need for a new range of standardized indicators based on reader (rather than author-facing) metrics.

**Design/methodology/approach** – This is a thought experiment, outlining the kinds of usage indicators that could be developed alongside the traditional ISI measures of impact, immediacy and obsolescence.

**Findings** – The time is ripe to develop a set of standardised measures of journal usage that are as easy to understand, and as universally accepted, as ISI’s current citation-based indicators. By linking article publication year to full text downloads, this article argues that very considerable value could be extracted from what, in many cases, is almost uninterpretable data.

**Practical implications** – Indicators in the form proposed could find a wide variety of applications, from helping librarians to assess the potential value-for-money of bundled journal deals, to helping policy-makers and scholarly communication researchers to better understand the dynamics of knowledge diffusion.

**Originality/value** – The development of standardized usage factors in the form suggested here would radically shift the centre of gravity in bibliometrics research from the author to the reader. This remains largely unexplored territory.

**Keywords** Serials, Quality, Measurement

**Paper type** Conceptual paper

## The need for usage-based metrics at the journal level

Citations tell us a lot about the scholarly communications system as seen from an author perspective. The citation indicators produced annually by ISI Thomson Scientific offer us insights into which journals are, on average, the most highly cited (impact factor), which journals researchers turn to first (immediacy index), and the long term value that authors ascribe to particular titles (cited half-life). These are important measures that are used extensively by publishers, librarians and policy makers to help them make sense of a complex and confusing world. But they only tell a small and increasingly less relevant part of the whole story.

The journals market devours something like seven billion dollars a year, a very high proportion of which comes from public sources, so there is a responsibility on all of us to make sure that the data we rely on is fit for purpose. The transition from a paper-based to a digital environment is forcing us to rethink the consumer landscape and how we measure the utility of journals. In today’s world of the digital library, however, there is a gaping hole in our understanding of scholarly communication. It is easy to forget that these two populations, authors and readers, are not the same (as

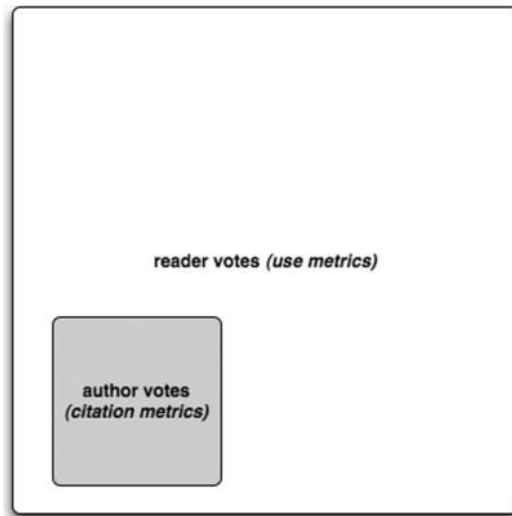


Figure 1 shows schematically). They have different needs and different requirements of the journals literature[1]. We are now at a point in the evolution of the journals system where we can systematically measure journal use (“votes by readers”) as well as journal citation (“votes by authors”). We could do this by exploiting the data trails (weblogs) that readers leave behind on publishers’ and aggregators’ websites.

In the hard copy era, it was simply not possible to view the scholarly communication system through the eyes of readers. The data were neither robust nor generally available and so we made do with small-scale, ambiguous, localised data in the form of library re-shelving statistics. This need no longer be the case since we can now directly measure usage in digital library environments. Not all readers write papers, many are practitioners, research users, policy makers and opinion formers, translating research into actionable outcomes, so they are an important constituency.

ISI indicators have always generated considerable interest among switched-on librarians, authors and publishers because they tell us a lot about researchers’ behaviour. A credible series of usage-based indicators would be of enormous interest to the scholarly community, from first-time authors choosing where to publish to international bodies evaluating the effectiveness of their grants programme. The juxtaposition of citation and usage indicators would add fresh new perspectives to our understanding of journal productivity in the round. They would also be credible. A survey of 5,513 senior researchers by CIBER (Rowlands and Nicholas, 2005) found that 69 per cent of respondents agreed with the proposition that downloads offer a good indication of the “usefulness of research” (a similar but slightly small proportion, 67 per cent, agreed that citations were a similarly useful indicator!).

Ask yourself a question. The last time you bought a digital camera or MP3 player, how many product attributes did you consider before making a decision? Almost



**Figure 1.**  
Author and readers:  
overlapping but distinct  
populations

---

certainly more than are currently available in a convenient form for discriminating between journals costing ten times as much. In the next section, we propose a set of usage metrics (“classic clones” that mimic ISI’s citation-based metrics[2]. We have concluded that COUNTER data do not offer a useful platform for the development of standardised usage metrics since they do not currently capture the publication year of any downloaded materials. In all of the cases, however, that CIBER has so far examined, publishers’ log data provide rich information at the individual article level, including an explicit year of publication.

### **Towards usage metrics at the journal level**

In this section we reflect on how publishers’ downloads might be used to mimic the four classic ISI citation metrics (total cites, impact factor, immediacy index and cited half-life) from a reader’s point-of-view. In each case, the new usage metric is calculated simply by replacing citations with downloads. The analogies are not perfect and remain to be tested. Unless otherwise specified, all the indicators that follow are based on a synchronous design and “download” refers to the full text article.

#### *Article download indicator*

Total cites is the simplest of the four journal citation measures, and it is helpful in illustrating the principle that downloads can profitably be transposed for citations to create useful new metrics. Calculation of the 2007 article download indicator: total number of articles downloaded from journal  $J$  during 2007, regardless of their age.

In neither world, citation nor download, is 100 per cent coverage a reasonable or even a desirable target to aim for. While individual publisher’s or aggregator’s log files will contain comprehensive use data for that platform, a journal may be mounted on a number of different systems. The nature and extent of this multiple platform issue is of course a potential major limitation.

#### *Article impact factor*

Garfield’s impact factor is the iconic journal metric. In its common sense meaning, the journal citation impact factor (CIF) represents the number of times an average “recent” article is cited in a given year, a “recent” article typically being defined as one or two years old. We can imagine similarly constructing a journal usage impact factor (UIF). Calculation of the 2007 usage impact factor: numerator (a) = article downloads from journal  $J$  during 2007, limited to publications in  $J$  from 2006 and 2005; denominator (b) = the number of articles published in  $J$  in 2006 and 2005;  $UIF_J = a/b$ .

Notes: in (a), citations to, or downloads from, any part of the journal are counted. In contrast, (b) is limited to particular “citeable” document types: articles, notes and reviews.

Logically, it should be possible to simply replace “citations” with a usage measure (full text downloads) in the standard impact factor calculation, if the data can be reliably extracted from the log files. Again, a common sense interpretation would be the answer to the question: “how many times is the average recent article in journal  $J$  downloaded within a given time window?”[3].

Citation journal impact factors can only be calculated for titles that ISI includes in its indexes. A further limitation is that the impact factor is incomplete: a journal may receive citations from journals not included in the ISI indexes (for example, many open

---

access journals) but these cannot be counted, so they do not feature in the Journal Citation Reports. Theoretically, there is no reason why we should not construct journal download impact actors (JDIFs) for any publisher and any journal; they do not need to have an ISI listing. This may be of considerable interest to publishing houses (even Elsevier's list includes a substantial number of non-ISI indexed titles) and the approach could in principle be extended to other publishing media, such as institutional and subject repositories, e-print servers and open access journals.

What are the most immediately obvious differences between CIFs and UIFs and to what extent are they comparable? Potentially, UIFs could offer comprehensive data on all uses if we were able to secure the publishers' logs for all their delivery platforms (CIFs necessarily offer a snapshot of citations received within an admittedly very large but incomplete universe). CIFs are medium- and platform-independent. It does not matter whether the author found the article in a print or electronic format: they might even write its details down on a napkin in a restaurant. Of course, UIFs only tell us about use in digital libraries and thus do not account for print-based use. This is an issue, especially for society journals with a large personal subscriber base and, of course, it does not work at all for print only titles. However, it is easy to overstate the problem. Two recent studies[4, 5] both find a strong positive correlation between the use of the printed and electronic versions of a journal. In other words, titles that are more heavily used in one format tend to be more heavily used in the other. As a very rough rule of thumb, online accesses now probably out-number print accesses by at least a factor of ten. This is good news: one of the obvious limitations of a usage metric based solely on downloads is that we would be open to the criticism that we were ignoring print users. The studies by Obst (2003) and by Wulff and Dixon (2004) are quite compelling and suggest that download metrics should be credibly be representative of all uses for a large majority of titles.

Usage impact factors would surely be extremely interesting to publishers: they would be very useful in informing pricing decisions and for evaluating the impact of marketing campaigns on subsequent usage. Librarians would find them much more meaningful for their purposes than the current measures based on citations and they could be used, for example, in evaluating the value-for-money aspects of Big Deals. Research funders and the science policy community would surely be interested as well since measures of actual use would speak directly to questions about knowledge transfer.

### *Usage immediacy index*

The immediacy index is a very simple measure in the citation world: how quickly very recent articles are cited after publication: it reveals which journals authors consider are really hot! Calculation of the 2007 usage immediacy index: numerator (a) = citations received (or articles downloaded) during 2007, limited to publications from 2007; denominator (b) = the number of articles published in 2007;  $UII_{2007} = a/b$ .

Notes: in (a), citations to, or downloads from, any part of the journal are counted. In contrast, (b) is limited to particular document types: articles, notes and reviews.

The common sense meaning of the parallel usage indicator, the usage immediacy index, would be what proportion of papers published in 2007, for example, are downloaded (rather than cited) during 2007. There is an issue of document type (or rather the lack of such explicit information in the log files) and we would have every

reason to expect that letters and review papers, for example, might differ considerably with respect to this particular metric.

### *Usage half-life*

The cited half-life indicates the perceived long-term value of a journal to the author community, expressed as the median age of the articles they cite in their own publications. Typically, the citation time curve shows a sharp initial increase (authors' attention tends to focus on more recent papers) then a long tail off. Calculation of the 2007 usage half-life: the median age of all articles in journal *J* downloaded during 2007, regardless of their age.

But authors' perceptions of the long-term value of the scholarly archive are likely to be different from those of other readers: students, medical practitioners and university teachers. We simply do not know for certain, although there are some very strong indications that these patterns are different[6], possibly as a result of a retrieval paradigm that prioritises topical relevance over date of publication.

In principle, this approach could be brought to bear on usage data since we can tag downloaded articles by publication year. This would be of great value for librarians and publishers and it raises the possibility, for instance, of differential pricing on the basis of the age and perceived value of archive materials. It might also be used to relegate materials to the electronic equivalent of the library stack.

There is, however, an issue. When an author writes a paper, he or she is able to go right back to the very first issue of a journal, perhaps to a paper published in the late nineteenth century[7] without restriction. In a digital library, the user may not have this luxury: they can only go back in time as far as the back file permits. This raises a serious issue of comparability between a cited half-life and a download half-life: the latter is constrained by the system and the goal posts might change significantly from one year to the next if the size of the back file were to be suddenly extended or reduced. In cases where all the users of the digital library have a complete run of a journal, back to Vol. 1 No. 1, then there is obviously no problem. In cases of disciplines with short half-lives and incomplete but long back files, the problem diminishes significantly.

### **Conclusions**

We think the time is ripe to develop a set of standardised measures of journal usage that are as easy to understand, and as universally accepted, as ISI's current citation-based indicators. Librarians and publishers are awash with mountains of usage data but the problem is one of data overload and little real comparability. By linking article publication year to full text downloads, we argue that very considerable value could be extracted from what, in many cases, is almost uninterpretable data. Thomson Scientific has established a powerful market presence for its citation metrics, but these are becoming less relevant as the journals system opens up to the digital consumer. Standardised usage metrics would be of enormous interest to publishers and would help them to understand their markets better and price their offerings on a more rational basis. Usage metrics would be far better attuned to the needs of librarians than ISI's current offerings and would, for instance, help them to negotiate better value-for-money from Big Deals and other offers. Usage metrics would transform research in science policy and scholarly



---

communication (big markets for ISI data) by opening up new questions about knowledge transfer. There are strong research indications that electronic and print uses are positively correlated. This, together with the fact that print access is in such dramatic decline, means that these metrics should win market confidence quickly. Unlike citation-based metrics, usage metrics could be created for any journal with an electronic version, including open access journals, and for a variety of delivery platforms, such as e-print servers and institutional repositories.

### Notes

1. Recent work by CIBER on a variety of publishers' logs has consistently failed to reveal any statistical associations between download frequencies and classic ISI impact factors. Indeed, there is no particular reason why there should be: articles are read for a wide variety of purposes, not simply for inclusion by authors in their reference lists.
2. An exhaustive search of the formal literature has only recovered one paper with a similar aim to this short communication: to create standardized measures of journal use (Darmoni *et al.*, 2002). This paper proposes a journal "reading factor" and this is generated in real time by counting hypertext clicks on an experimental system. Darmoni's reading factor does not attempt to mimic ISI's journal impact factor (as with COUNTER's data, there is no way of establishing the date of publication) and it is hardly conceivable that this approach could be scaled up much beyond a single library.
3. A very strict definition of the journal impact factor would address the so-called numerator/denominator problem by limiting the items in the numerator to ISI's "citable units": research articles, reviews and brief communications or notes (Amin and Mabe, 2000). It seems unlikely that we could distinguish whether a particular download was a book review or a major research article without considerable additional costs being invoked. This assertion has not been tested, however, and it may be possible to automatically match log records with the ISI database to establish document types at reasonable cost.
4. Obst (2003) found good correlations for print and electronic usage for Academic Press, Blackwell, Elsevier, HighWire and Springer titles, averaging around 0.67. He also found that the rankings of the top seven journals were identical for both print and electronic use.
5. Similarly, Wulff and Dixon (2004) found strong, significant correlations (in the range 0.55 to 0.85 with  $p < 0.01$ ) between the use of print and digital copies for the Ovid, Science Direct and IDEAL platforms.
6. A study by Tsay (1999) compared the obsolescence characteristics of journal citations and print journal readings (using re-shelving data) for a carefully controlled journal set of biomedical titles and found that the decay patterns were significantly different for citations and readings.
7. This would be the exception rather than the norm in the sciences, although it is conventional for authors in surgery to cite the original paper that introduced a particular surgical technique.

### References

- Amin, M. and Mabe, M. (2000), "Impact factors: use and abuse", *Perspectives in Publishing*, No. 1, pp. 1-6.
- Darmoni, S.J., Roussel, F. and Benichou, J. (2002), "Reading factor: a new bibliometric criterion for managing digital libraries", *Journal of the Medical Library Association*, Vol. 90 No. 3, pp. 323-7.

- Obst, O. (2003), "Patterns and costs of printed and online journal usage", *Health Information and Libraries Journal*, Vol. 20 No. 1, pp. 22-32.
- Rowlands, I. and Nicholas, D. (2005), "New journal publishing models: the 2005 CIBER survey of journal author behaviour and attitudes", *Aslib Proceedings*, Vol. 57 No. 6, pp. 481-97.
- Tsay, M.Y. (1999), "Library journal use and citation age in medical science", *Journal of Documentation*, Vol. 55 No. 5, pp. 543-55.
- Wulff, J.L. and Dixon, N.D. (2004), "Quality markers and use of electronic journals in an academic health sciences library", *Journal of the Medical Library Association*, Vol. 92 No. 3, pp. 315-22.



**This article has been cited by:**

1. Dion Curry, Steven Van de Walle. 2018. A Bibliometrics Approach to Understanding Conceptual Breadth, Depth and Development: The Case of New Public Management. *Political Studies Review* 16:2, 113-124. [[Crossref](#)]
2. Bikun Chen. 2018. Usage pattern comparison of the same scholarly articles between Web of Science (WoS) and Springer. *Scientometrics* 115:1, 519-537. [[Crossref](#)]
3. Negin Salimi. 2017. Quality assessment of scientific outputs using the BWM. *Scientometrics* 112:1, 195-213. [[Crossref](#)]
4. Liwen Vaughan, Juan Tang, Rongbin Yang. 2017. Investigating disciplinary differences in the relationships between citations and downloads. *Scientometrics* 111:3, 1533-1545. [[Crossref](#)]
5. Muhammad Salman Khan, Muhammad Younas. 2017. Analyzing readers behavior in downloading articles from IEEE digital library: a study of two selected journals in the field of education. *Scientometrics* 110:3, 1523-1537. [[Crossref](#)]
6. SingsonMangkhollen, Mangkhollen Singson, ThiyagarajanS., S. Thiyagarajan, LeeladharanM., M. Leeladharan. 2016. Relationship between electronic journal downloads and citations in library consortia. *Library Review* 65:6/7, 429-444. [[Abstract](#)] [[Full Text](#)] [[PDF](#)]
7. . References 131-153. [[Crossref](#)]
8. Kim Holmberg. The Past 9-54. [[Crossref](#)]
9. Ehsan Mohammadi, Mike Thelwall, Stefanie Haustein, Vincent Larivière. 2015. Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology* 66:9, 1832-1846. [[Crossref](#)]
10. Wolfgang Glänzel, Juan Gorraiz. 2015. Usage metrics versus altmetrics: confusing terminology?. *Scientometrics* 102:3, 2161-2164. [[Crossref](#)]
11. Peter Kraker, Christian Schlögl, Kris Jack, Stefanie Lindstaedt. 2015. Visualization of co-readership patterns from an online reference management system. *Journal of Informetrics* 9:1, 169-182. [[Crossref](#)]
12. Stefanie Haustein, Isabella Peters, Judit Bar-Ilan, Jason Priem, Hadas Shema, Jens Terliesner. 2014. Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics* 101:2, 1145-1163. [[Crossref](#)]
13. Juan Gorraiz, Christian Gumpenberger, Christian Schlögl. 2014. Usage versus citation behaviours in four subject areas. *Scientometrics* 101:2, 1077-1095. [[Crossref](#)]
14. Zohreh Zahedi, Rodrigo Costas, Paul Wouters. 2014. How well developed are altmetrics? A cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications. *Scientometrics* 101:2, 1491-1513. [[Crossref](#)]
15. Ehsan Mohammadi, Mike Thelwall. 2014. Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology* 65:8, 1627-1638. [[Crossref](#)]
16. Stefanie Haustein, Vincent Larivière. 2014. A multidimensional analysis of Aslib proceedings – using everything but the impact factor. *Aslib Journal of Information Management* 66:4, 358-380. [[Abstract](#)] [[Full Text](#)] [[PDF](#)]
17. Daniel Torres-Salinas, Evaristo Jiménez-Contreras, Nicolás Robinson-García. 2014. Tendencias en mapas de la ciencia: co-uso de información científica como reflejo de los intereses de los investigadores. *El Profesional de la Informacion* 23:3, 253-258. [[Crossref](#)]

18. Zahed Bigdeli, Morteza Kokabi, Gholam Reza Rajabi, Ali Gazni. 2013. Patterns of authors' information scattering: towards a causal explanation of information scattering from a scholarly information-seeking behavior perspective. *Scientometrics* **96**:1, 103-131. [[Crossref](#)]
19. Mike Thelwall, Stefanie Haustein, Vincent Larivière, Cassidy R. Sugimoto. 2013. Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE* **8**:5, e64841. [[Crossref](#)]
20. Stefanie Haustein, Tobias Siebenlist. 2011. Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics* . [[Crossref](#)]
21. Christian Schloegl, Juan Gorraiz. 2011. Global usage versus global citation metrics: The case of pharmacology journals. *Journal of the American Society for Information Science and Technology* **62**:1, 161-170. [[Crossref](#)]
22. Valérie Durieux, Pierre Alain Gevenois. 2010. Bibliometric Indicators: Quality Measurements of Scientific Publication. *Radiology* **255**:2, 342-351. [[Crossref](#)]
23. Christian Schloegl, Juan Gorraiz. 2010. Comparison of citation and usage indicators: the case of oncology journals. *Scientometrics* **82**:3, 567-580. [[Crossref](#)]
24. Michael J. Kurtz, Johan Bollen. 2010. Usage bibliometrics. *Annual Review of Information Science and Technology* **44**:1, 1-64. [[Crossref](#)]
25. Mike McGrath. 2007. Interlending and document supply: a review of the recent literature: 61. *Interlending & Document Supply* **35**:4, 211-218. [[Abstract](#)] [[Full Text](#)] [[PDF](#)]