The use and abuse of usage measures

Updated from a paper presented at the UKSG annual conference, Torquay, April 2008[1]

This paper looks at the extent to which the COUNTER guidelines on usage data have met their goals for recording and reporting usage data in a consistent, credible and compatible way. It compares the proposed Usage Factor with the Citation Impact Factor and highlights key differences between the act of downloading and the act of citation. Drawing on the experience of the Impact Factor, it then speculates on how measuring the literature can change the literature and how this Observer Effect might impact usage patterns.

IAN BANNERMAN
Managing Director, Journals
Taylor & Francis

Introduction

It is easy to be seduced by the promise of using online usage data as a measure of quality and value, because we can now measure online use far better than we could ever measure print use, and because of concerns that citation measures have become too dominant in this area. As a result, all those involved in scholarly communication – authors, editors, librarians, publishers and intermediaries – have an appetite for usage data and are keen to draw conclusions from them, but it is important to understand some of the limitations of usage measures and to consider the distortions and biases that they might bring into the equation. When Eugene Garfield introduced the Journal Citation Impact Factor he unleashed a chain of unintended consequences that continue to shape the way that research is published and valued. In the proposed development of a Usage Factor we can learn a good deal from our experiences with the Citation Impact Factor and hopefully avoid some of the unintended consequences that a new measure/ranking might introduce.

COUNTER and the Usage Factor

The COUNTER initiative was launched in March 2002, and describes itself as 'an international initiative serving librarians, publishers and intermediaries by setting standards that facilitate the recording and reporting of online usage statistics in a consistent, credible and compatible way.'[2] The key words here are 'consistent', 'credible' and 'compatible' and without COUNTER we would be unable to combine or compare usage data, or trust that they had been compiled in a way that eliminates some of the more obvious reporting errors. The COUNTER team have made considerable progress in creating standard reports and in defining what counts as a download and what should be excluded. There is more to be done, however, particularly in the detection and reporting of robot activity. Consistency, credibility and compatibility remain excellent goals for COUNTER and are the essential foundation for any usage-based measure or ranking.

The number of downloads from a journal will be affected by the number of articles that are available online. All other things being equal, a journal with many articles would be used more than a journal with fewer articles. Eugene Garfield was faced with a similar problem in comparing citations across journals. His simple solution was the creation of the Citation Impact Factor, which is a measure of the average number of citations per

article[3]. The 2007 Citation Impact Factor, for example, is defined as:

$$\frac{\text{Total cites in 2007 for items published during 2005/6}}{\text{Total items published during 2005/6}}$$

Johan Bolen and Herbert Van de Sompel used the same logic to define a Usage Impact Factor:

$$\frac{\text{Total usage over period 'x' of items published during period 'y'}}{\text{Total items published online during period 'y'}}$$

They then compared it with Citation Impact Factor using data from California State University[4]. This was picked up by the COUNTER group and in October 2006 COUNTER director Peter Shepherd interviewed 7 authors, 9 librarians and 13 publishers to gauge enthusiasm for the creation of a Usage Factor ranking for journals. This was followed up by a web survey of 155 librarians and 1,400 authors. The results were published by the UK Serials Group (UKSG) in June 2007[5], and in 2008 UKSG plan to issue an invitation to tender 'to investigate and test the feasibility of developing a new metric, the Journal Usage Factor, based on COUNTER usage data.'

Algebraically, the proposed Usage Factor is very similar to the Thomson Scientific Impact Factor. Both of them attempt to correct for the size of the journal, to create a measure of the average activity per article. Despite this superficial similarity, it is important to recognize that there are very clear differences between citations and downloads. If citations can be characterized as a measure of prestige, then downloads are a measure of popularity. Citations and downloads are done by different people at different times and for different reasons. At the moment, decisions about careers, funding and the focus of research are not influenced by usage data, but they are certainly influenced by citation data.

When developing the concept of a Usage Factor, there are two implicit assumptions. First, that usage data is consistent, credible and compatible, as COUNTER intends it to be. Secondly, that the Usage Factor would be a meaningful indicator of something else, such as quality or value, otherwise it would be a pointless exercise.

Consistent, credible and compatible usage data

One of the biggest threats to the integrity of usage data comes from the activity of robots, leading one publisher to claim that one out of every two downloads is made by machine or by mistake.[6] If that really is the case, and the spurious data are not filtered out, then there can be little confidence in a Usage Factor. Legitimate crawling by search engines and text-mining software are vital in maximising the discoverability and utility of content, but it is important that this activity does not distort the reporting of usage. Release 3 of the COUNTER guidelines makes a first attempt at getting to grips with the issue of robots and 'pre-fetching' (loading content into a cache in order to optimise the speed of downloads)[2]. At the heart of these guidelines is the creation of a list of known robots that publishers should then filter out. This is a helpful step, but it does not solve the problem. In a subscription-controlled environment most robots will fail anyway because they will be blocked by access controls (although this is not the case for open access journals). Those that are allowed in (e.g. Google) can be easily excluded from reports. The real concern, however, is the 'home-built' robots that operate within the IP ranges of legitimate customers and it is these that currently cause the most distortion.

Robot activity is easier to spot when it goes 'wrong'. For example, a robot within a Russian institution that was presumably intended to harvest a range of our articles got stuck on retrieving just one but did so 6,372 times. Fortunately, this would not have registered on a COUNTER report because of the 'double-click' requirement to filter out downloads of the same article made within 30 seconds. Other examples have got through to COUNTER reports, however, including a robot from Korea that harvested all of the articles from an economics journal and was only detected because it ran 58 times. The difficulty here is that some of this activity is only discovered after the reports have been produced and analysed, so more guidance is required on the procedure for correcting COUNTER reports retrospectively.

To date, I am not aware of a robot that has been designed to deliberately inflate usage statistics but if the Usage Factor were to become an influential metric then there might be an incentive to do so. The 'best' robots are very good at mimicking human activity but some humans are mimicking robots – Philip M Davis from Cornell reports on LibLicense[7] that "I've successfully downloaded my own article thousands of times from a reputable publisher without setting off any alarms." Some

activity defies explanation, and we have recently picked up an example of a niche article on fish in Mauritius that has been downloaded over 1,000 times from various institutions but has never been cited.

The Usage Factor as a meaningful indicator?

There is clearly a good deal of 'noise' in the system, caused by both robot and human activity, which is inflating article downloads and potentially obscuring genuine readership. COUNTER is making good progress in minimizing this but the 'noise' will always remain. More fundamentally, there is no way of recording the intention behind a download, whereas the intention behind a citation is recorded in the literature itself.

At the end of February 2008, the web analytics firm comScore published a report that showed a 7% decline in the click-through rate for Google ads. This contributed to a 4% fall in Google's share price. comScore then published a clarification, stating "the evidence suggests that the softness in Google's paid click metrics is primarily a result of Google's own quality initiatives that result in a reduction in the number of paid listings"[8]. It appears that by providing a more targeted service to their users, Google had reduced their click metrics and damaged their share price. If the widespread use of a Usage Factor had a similar effect on online journals, there would be a disincentive to direct users to relevant content. This would be a harmful unintended consequence that we should seek to avoid.

Our experience at Taylor & Francis is that overall usage levels are strongly influenced by changes in interface design. Quick and easy search-and-browse functions and reducing the number of clicks to content are simple measures that all publishers deploy in order to build usage. These interface effects were explored by Davis and Price[9], who found significant variation between publisher interfaces in the ratio of PDF to HTML articles. Those interfaces that encourage the user to view both the PDF and the HTML version will record higher usage because both versions 'count'. Davis and Price found that this ratio of PDF:HTML downloads varied from 1 to 19.8, suggesting that interface alone (regardless of article quality or value) can have a significant impact on overall usage.

David Nicholas and his colleagues from CIBER have done a lot of work on user behaviour and identified patterns that cast further doubt on the reliability of usage data as a meaningful indicator of quality or value. The skimming of the literature that they describe would suggest that many articles are viewed but few are read[10]. Herbert van de Sompel and his colleagues from Los Alamos are looking at usage data and citation data under the MESUR project and highlighting some of the differences between citation as a prestige measure and usage as a popularity measure[11]. As these studies progress we will learn more about the factors that influence usage data and the correlation (or lack of it) with other metrics.

The Observer Effect

The Observer Effect refers to changes that the act of observing will make on the phenomenon being observed. The CIBER group and the MESUR project are fortunate to be working in a relatively pristine environment, where they can study the use of the literature without influencing it. Publishers are working hard to drive up online usage, through improved sales, marketing and online functionality, but the editorial content of the journal has been largely unaffected by the publishers' need to build usage. Citation rankings, by contrast, have become so pervasive that they have begun to exert an influence on the way that research is conducted and published. Most of these are well known and understood by those that publish and consume citation rankings. The list below is for illustration only and certainly not intended as a 'to-do' list for editors:

Some of the ways in which the Citation Impact Factor has influenced the academic literature:

- self-citing the journal in other articles and editorials
- alerting authors to content they 'should' cite
- seeking out prolific, high quality authors – who will self-cite
- building editorial boards that will attract citations
- publishing themed issues with prestigious guest editors
- publishing the most citable papers early in the year

- keeping review times short so citations don't miss the 2-year impact factor window
- targeting topical areas rather than long-term studies
- publishing review articles
- publishing news, letters, obituaries, book reviews, editorials (that get cited without counting as citable articles).

Such activity (not all of it harmful) goes on because the academic community is complicit in the need to drive up citations. Careers, research funding and even salaries have all been linked to citation measures but not (yet) to usage. Eugene Garfield, the creator of the Impact Factor, has said that "In 1955 it did not occur to me that 'impact' would become so controversial."[12] The lesson to be learned from the Impact Factor is that by measuring the literature, we change the literature. This has led to claims from Richard Monastersky (2005)[13] that the impact factor is the "number that's devouring science".

If we were to introduce the Usage Factor and it, too, were to become a significant factor in determining careers, research funding and even salaries, then The Observer Effects might be even more extreme:

Some of the ways in which the proposed Usage Factor might influence the academic literature:

- getting your friends, your dog and your mother to download articles … or writing a bot to do it for them
- leaving usage data unfiltered … or worse
- publishing for students, not for researchers
- 'Sexing up' title and key words
- putting the HTML in the way of the PDF
- using the abstract to tease, not to inform
- stopping printed journals
- encouraging online coursepacks, discouraging printed ones
- blogging it, tagging it, posting it, to extremes
- broadcasting metadata but keeping articles where they are counted – not in open access repositories!

There are important differences between the act of citing and the act of downloading, which I believe make the Usage Factor even more prone to these Observer Effects. Attempts to manipulate the Impact Factor leave a trail in the literature which can always be re-examined by anyone at any time;

attempts to influence the Usage Factor may leave a mark in the usage logs but these are hard to interpret and not available for public scrutiny. The act of downloading is often meaningless, done by mistake, done by a robot, done because the interface encouraged you to do something that you might not have intended. Downloading requires little investment and is practically anonymous; citation is usually meaningful and requires significant investment of time, effort and reputation.

Recommendations

1. Beware of over-interpreting usage data. I am not opposed to the development of the Usage Factor, just cautious about how it might be used and wary of the over-interpretation of usage data. James Pringle (2008)[14] warns about the over-interpretation of citation data and provides two 'golden rules' that apply equally well to usage data:

   - first, "consider whether the available data can address the question"
   - second, "compare like with like".

   The COUNTER team are to be commended for the careful consideration that they are giving to these issues and their commitment to consultation along the way.

2. More work needs to be done on the detection, blocking and filtering of robots and other aberrant downloads, without blocking the legitimate use of authorized crawling and text mining. COUNTER is beginning to address this but we need stronger guidelines on how to monitor robots within subscribing IP ranges and what actions to take when such use is discovered after usage reports have been published.

3. More research is needed into the factors that influence usage statistics. In particular, the MESUR studies at Los Alamos and the work of the CIBER team at University College London are giving us fresh insight into this area, but this is a young field of research and by definition there are no long-term longitudinal studies on how usage patterns change over time. This research is vital if we are to place any value on the Usage Factor as a meaningful and reliable metric.

References

1. Bannerman, I, Presentation given at the UKSG annual conference, Torquay, April 2008: http://www.uksg.org/sites/uksg.org/files/ian_bannerman.pps (Accessed 6 June 2008)

2. COUNTER website: http://www.projectcounter.org (Accessed 6 June 2008)

3. Thomson Scientific Impact Factor: http://scientific.thomson.com/free/essays/journal citationreports/impactfactor/ (Accessed 6 June 2008)

4. Bollen, J and Van de Sompel, H, Usage Impact Factor: the effects of sample characteristics on usage-based impact metrics, 2006:arXiv:cs. DL/0610154 v2

5. UKSG Usage Factor: http://www.uksg.org/usagefactors (Accessed 6 June 2008)

6. Berkley Electronic Press http://www.bepress.com/download_counts.html (Accessed 6 June 2008)

7. Davis, P M, LibLicense, 2007: http://www.library.yale.edu/~llicense/ListArchives/0710/msg00040.html (Accessed 6 June 2008)

8. Magid Abraham, M, Why Google's surprising paid click data are less surprising, 2008: http://www.comscore.com/blog/2008/02/why_googles_surprising_paid_click_data_are_less_surprising.html (Accessed 6 June 2008)

9. Davis, P M, and Price, J S, eJournal interface can influence usage statistics: implications for libraries, publishers, and Project COUNTER, *JASIST*, 2006, 57(9),1243–1248.

10. Nicholas, D, The information-seeking behaviour of the virtual scholar: from use to users, *Serials*, 2008, 21(2), 89–92. doi: http://dx.doi.org/10.1629/2189 . See also presentation given at the UKSG annual conference, Torquay, April 2008: http://www.uksg.org/sites/uksg.org/files/david_nicholas.pps (Accessed 6 June 2008) on which article is based.

11. Van de Sompel, H, presentation given at the UKSG annual conference, Torquay, April 2008: http://www.uksg.org/sites/uksg.org/files/herbert_van_de_sompel_(MESUR).pdf (Accessed 6 June 2008)

12 Garfield, E, The Agony and the Ecstasy – the History and the Meaning of the Journal Impact Factor. Presented at the International Congress on Peer Review and Biomedical Publication, Chicago 2005: www.pitt.edu/~super1/21011-22001/21231.ppt (Accessed 6 June 2008)

13. Monastersky, R, The Number That's Devouring Science, *Chronicle of Higher Education*, October 2005

14. Pringle, J, Trends in the use of ISI citation databases for evaluation, *Learned Publishing*, 21(2), 85–91. doi: 10.1087/095315108X288901

Article © Ian Bannerman

■ **Ian Bannerman**
Managing Director, Journals
Taylor & Francis
4 Park Square
Milton Park
Oxfordshire
OX14 4RN
e-mail: ian.bannerman@informa.com