

Electronic Journal Usage Statistics: Present Practice and Future Progress

Tony Kidd

Head of Serials/Document Delivery, Glasgow University Library, UK

Abstract

Comparable and standardised e-journal statistics are desperately needed by library staff for a number of purposes, including collection management, collection promotion, budget allocation, and budget negotiation. This paper reviews the requirement for usage statistics, and then moves on to consider some of the efforts that have been taking place over recent years to promote and implement standards in this area, concentrating on the international work in progress under the auspices of Project COUNTER.

Introduction

The majority of papers given at this satellite conference have discussed performance measurement relating to evaluation of library activities, using statistics gathered within the library itself, in order to monitor, and more especially improve, the effectiveness of services undertaken for library users, and to ensure that library staff resources, for example, are deployed in an optimum way. Measurement of electronic journal usage is primarily (though not entirely) dependent, however, on information supplied from outside the library. This makes standardisation perhaps even more of an issue than it is for other aspects of library statistics, and I shall spend a good part of my paper describing Project COUNTER, the current initiative to encourage that standardisation. Before that, however, I shall briefly review the reasons why electronic journal usage statistics are so important to have, and look at some of the existing statistics available from publishers and others.

It is no news to any librarian here that electronic journals have become much more widespread, and a much more important resource, over recent years. The figures will vary from library to library, and from country to country, but the trend is, I imagine, universal. There are those who predict the almost complete disappearance

of print journals. It is not necessary to go that far to note that there are already many libraries with more titles available online than in print. Even some large research libraries in the UK, for example, are cancelling print journals in favour of the online versions (despite the built-in tax disincentives, particularly in the UK, where 17.5% Value Added Tax/Sales Tax is levied on electronic products, with zero taxation on print subscriptions). That trend is likely to accelerate sharply as really credible and robust archiving methods are implemented, guaranteeing access into the future both for the community in general, and for individual libraries that have subscribed to individual titles.

Use of e-journals

Libraries spend a large proportion of their budgets on serials. The most recent full-year Glasgow figures are shown in Table 1. Such a level of expenditure requires scrutiny to ensure that it is spent to best advantage, in terms of the value obtained by staff and students. The best measure of value is probably usage, if that can be accurately ascertained. It has always been very difficult to measure the use of print journals, at least in open access libraries, given that the majority of journal use is not reflected in loans. Practices such as asking users to annotate labels when reading issues, or not to reshelve issues or volumes; or attaching a

thread which breaks on handling of an issue; or even measuring the layers of dust on volumes – are all open to abuse or uncertainty of a fairly radical nature. The advent of electronic journals at least in theory allows some of these difficulties to disappear, although the practice is by no means so simple, as I shall discuss shortly.

Table 1 Expenditure on e-journals

2000/01 Glasgow expenditure (SCONUL)	
Print serials	£473,000
Online serials	£70,000
Print+online serials	£715,000
Online databases	£274,000
Print etc books	£709,000
Online books	£87,000
Document delivery	£86,000
Total	£2,414,000

As well as indicating whether the library is receiving ‘value for money’ for particular journal subscriptions, usage statistics are also valuable for marketing purposes. Low use, however measured, can be an indication of low value and therefore a pointer towards cancellation, but it can also be an encouragement to engage in publicity exercises on behalf of particular journals, or, more likely, with particular academic departments, if subject or reference librarian staff, or academics, feel that these journals ‘ought’ to be being consulted more. The number of relevant users needs also to be remembered: a specialist humanities journal may be essential for two or three staff in a particular department, but is never going to have usage figures remotely resembling those for *Nature* or *Science*. Analysing electronic journal usage statistics is not a mechanical exercise, and requires as much interpretation as the analysis of any other type of library statistics. Liz Hart’s paper on benchmarking is also relevant here. Usage data available from other ‘comparator’

libraries can be very useful, giving an indication of whether usage in your library is at, below, or above the level that might be expected, although of course again analysis has to take account of the circumstances of individual institutions. In most cases, in fact, high use rather than low use is recorded, and this in itself can be welcome ammunition in persuading funding committees of the value of electronic information and access, and encouraging funding to be a little more generous (or a little less ungenerous) than it might otherwise have been.

Another potential function of electronic journal statistics is to help with budget allocation, and I’m interested to hear Wanda Dole’s paper on the broader subject of statistics and budget allocation. At Glasgow we have a rather complicated formula for allocating the budget among the different university faculties, but electronic journal use is included, contributing about 5% of the weighting (incidentally, we give about the same, or a little less, weighting to print journal use, however shakily measured: such statistics as we have indicate that usage of our electronic journals slightly exceeds usage of the print titles). Although most of our electronic journals usage information comes externally, from publishers, etc, they cannot provide us with information on differential faculty by faculty use, as all, or almost all, our university network traffic goes through local caches. We therefore rely on figures collected on a sample basis (hoping that the samples are a reasonable reflection of reality) by the University’s Computing Service, as indicated in Table 2. These figures suffer from various drawbacks. For example, you can see that ‘other’ is the biggest category: this represents accesses from computer clusters in the Library and other non-faculty based locations. The tacit assumption, which may or may not be correct, is that this is split along the same proportions as the more easily identifiable use. Another problem is that every click on a publisher’s website is being measured here. There is no possibility

of discriminating between views of the complete full-text article, an abstract, table of contents, list of issues, a journal homepage, etc, for which we have to rely on the publisher's statistics. Finally, although the great majority of traffic goes through the university caches, there are one or two cases where this may not be so, and this may adversely affect the statistics for particular departments/faculties. Despite these caveats, we still believe that worthwhile data are compiled, which we do employ in the real world of budget allocation.

Table 2 Use of e-journals

Computing Service statistics at Glasgow UL	
Clin Medicine	26.7%
Biomed & Life Sci	22.3%
Physical Sci	5.2%
Vet Medicine	5.0%
Engineering	3.5%
Soc Sci	2.0%
Dental	1.8%
Law/Financial	1.8%
Arts	1.4%
Comp/Math Sci	1.1%
Education	0.1%
Other	28.5%

The current position

So, we do believe that electronic journal usage statistics are important, serving a number of different purposes. How useful are the current statistics we receive? As already indicated, I am concentrating here on external statistics (with the exception of our local statistics just mentioned), although in our early days with electronic journals we did collect more internal information. We were able to do that because we created a separate web page for each journal title as part of our access procedures, and an assumption, probably not justified even in those early stages, was that the majority of users would seek access through the Library web pages. It soon became too burdensome to create a page for each title, so that there is

no possibility of collecting local per title web page hit information. (What happens now is that a standard catalogue record is created, although this is also linked to from a simple alphabetical and broad subject listing, so that users have the option of either going via the OPAC or linking to the OPAC from a web listing). In addition, I'm sure (though know of no measurement of this) that a large proportion of the use of electronic journals will not go via the library pages at all, but rely on researchers local knowledge and bookmarking, or of course direct access via various database services, or linking mechanisms such as CrossRef and SFX.

As far as external statistics are concerned, the position is very much better than it was even just a year or two ago. Most large or indeed medium-sized publishers provide usage statistics of some kind, as do most of the consolidators and aggregators of electronic journal services. Competitive pressures have led to the more widespread provision of this information, encouraged by demands from library customers, both individual libraries and consortia, and aided by groups such as ICOLC (International Coalition of Library Consortia), whose revised edition of guidelines was promulgated last year (ICOLC, 2001) – many publishers declare that they are following ICOLC guidelines – or the UK's NESLI (National Electronic Site Licence Initiative), which along with many other consortia has tried to insist that usage statistics are provided as part of any consortium deal.

The data provided vary considerably, however, from publisher to publisher. Without any standards, publishers measure different things, have varying definitions for what they do measure, and present statistics to libraries through a number of different routes and in different ways. For libraries, this has a number of implications. Although it is usually possible to compare the use of individual titles from a given publisher or aggregator, it is almost impossible to

combine these statistics, so that, for example, an overview can be obtained of journals in a particular subject area, or funded by a particular department. Even where the same, or at least similar, things are being measured by different publishers, and theoretically they could be combined within a single table by library staff, the different methods of presentation make combination such a time-consuming task that, in practice, it is not undertaken.

There is also the question of combining figures for use of the same title via different services – including perhaps the publisher's own site, and one or more intermediaries or aggregators. For some purposes, the library will wish to keep separate figures for different services – does the use via that route justify a payment for the intermediary service? Should library staff be encouraging users in one direction rather than another? While for other purposes, a combined figure is important – what overall benefit are we getting for our subscription to a title? Another concern is the route users are following to reach information. Is there still a substantial amount of 'browsing' i.e. users going to the websites of particular journals (either directly or via the library catalogue or listings, or perhaps via a TOC table of contents service, such as the British Library's ZETOC service), and checking the contents of particular issues for articles of interest to them; or is there rather more what might be called 'searching', checking bibliographic databases for articles on a particular topic, and going direct from the database to the full text (if possible) of individual articles wherever they might be published? A third possibility arises from the growing availability of links, via CrossRef and/or open URLs, allowing researchers to move from one article to another via references or citations. In theory, usage statistics should be able to provide at least some information along these lines (although the practice is rather more tricky) to allow libraries to gain more of an insight into how their patrons use and access

information, and to adjust their services accordingly.

Future developments

Having stated the problem, I now want to move on to what we hope might be at least an initial solution. Nearly two years ago now, JISC, the Joint Information Systems Committee of the UK Higher Education Funding Councils, the UK Publishers Association, and the UK-based Association of Learned and Professional Society Publishers, ALPSP, set up a Usage Statistics Working Party to look at solutions for the problems I have been outlining, principally by generating standards in this area for adoption by the suppliers of electronic journals (publishers, intermediaries, etc). It very quickly became apparent that, first of all, there was absolutely no point in trying to get standards adopted just in the UK, or just by UK publishers. Given the international nature of information dissemination and the publishing industry, standards had to be promulgated and adopted internationally, or not at all. It also became apparent that there was a lot of work going on by different bodies on different aspects of this problem. A prime example is the US Association of Research Libraries' e-metrics project which Julia Blixrud is going to describe for us in the next paper, so I won't say anything further on that just now, other than to say that I understand that its main emphasis is on statistics gathering within the library, rather than obtaining comparable external statistics. ICOLC has already been mentioned, and other bodies with an interest include the US National Commission on Libraries and Information Science (NCLIS), NISO and ISO.

Although much was happening in this area, it did not appear to us that the particular concern of trying to get suppliers to collect, at source, consistent and compatible statistics, was specifically being addressed by any group, indeed the ARL e-metrics project originally concluded that "*it is*

largely impossible to compare data across vendors, and we recommend that comparison be limited to data from the same vendors". (Shim et al, 2001)

There was still a feeling that it was worth trying to achieve the impossible, and this has been borne out by the expansion of the original Usage Statistics Working Group into a very much wider and more representative international initiative, entitled COUNTER, Counting Online Usage of NeTworked Electronic Resources, with a website at <http://www.projectcounter.org>. COUNTER has been endorsed by a wide range of relevant bodies (including some financial support from the majority), namely:

- AAP, Association of American Publishers
- ALPSP, The Association of Learned and Professional Society Publishers
- ARL, Association of Research Libraries
- ASA, Association of Subscription Agents and Intermediaries
- BIC/EDIeUR
- JISC, Joint Information Systems Committee
- NCLIS, National Commission on Libraries and Information Science
- NISO, National Information Standards Organization
- PA, The Publishers Association
- STM, International Association of Scientific, Technical & Medical Publishers
- UKSG, United Kingdom Serials Group

In addition, many individual publishers and intermediaries have already provided some financial support, and/or are actively involved in the various working parties and groups established by COUNTER.

Project COUNTER

Since its inception, COUNTER and its predecessor group has been chaired by Richard Gedye of Oxford University Press. Following a very well-attended meeting of many of the stakeholders from the

publishing and library worlds in London last June, it was realised in due course that voluntary activity by those interested, although vital for eventual success, was insufficient to take COUNTER through to eventual implementation, and since March this year Peter Shepherd has been acting as COUNTER Project Director. There is now a Steering Group of around thirty members, including a number of individual librarians from Europe and North America representing academic and corporate libraries, together with representatives from most of the bodies listed above and from a number of individual publishers and intermediaries, including American Chemical Society, Blackwell, CABI, EBSCO Publishing, Elsevier, Ingenta, Institute of Physics Publishing, Nature Publishing Group, Oxford UP, and Wiley.

COUNTER's aim is to have Release 1 of a Code of Practice issued and ready for use by the end of this year, compatible with ICOLC's Guidelines and with other developments. Different task forces have been looking at different aspects of this, and you may have been aware, for example, of the Online Librarian Survey which has been available for the last couple of months or so on the COUNTER website. If you haven't yet completed this, you should do so today or tomorrow, if you can, as it has a closing date of 15 August 2002. This survey is checking on the content, format and frequency of Usage Reports that will be most useful to librarians, and covers both journals and databases (although I have concentrated on journals today, COUNTER also includes usage statistics for databases). When I last heard, well over 500 responses had been received from all over the world, and from all types of libraries, suggesting a real interest in the library community in what COUNTER is trying to achieve.

Another task force is looking at the problems of definition – what exactly is a fulltext article download, what about distinctions between HTML and PDF, etc:

NISO's recently issued Data Dictionary (NISO, 2002) for its Z39.7-2002 standard on Metrics & Statistics for Libraries and Information Providers will be useful here. There is also a task force considering technical measurement matters – for example, how to decide exactly what to do about double-clicks: how long a gap does there have to be for a new click to count as a new use, and to what extent is this complicated by the operation of proxy servers and local caches? In one sense, the exact solution to these various conundrums is not vital: what does matter is that each publisher or intermediary should, as far as possible, be measuring in the same way, in order that statistics are compatible and credible.

So, to recap the aim is to get a Code of Practice issued by the end of the year. The Code is likely to be revised as experience is gained, although in ways compatible with previous versions. It will also likely have three levels of compliance, minimum, intermediate and highest, so that publishers are not inhibited from participating initially, but also encouraged to develop their statistics. It is envisaged that publishers will

be able to say that they comply with particular levels of the COUNTER Code of Practice, and that there will be some kind of auditing service to monitor compliance, and build up confidence in the system.

The essential point, of course, is that the Code is widely adopted and well promoted and marketed to publishers, librarians and consortia. Given the investment, in time, sponsorship and money, that most relevant groups have put into COUNTER already, and plans to continue development after this year, we believe the omens are good and that the great majority in the publishing industry will be willing and able to comply, and that libraries will soon be able to benefit from compatible and comparable usage statistics.

I hope many of you will be present in Glasgow next week for the IFLA conference proper. After a slightly slow start, I think the atmosphere in Glasgow is now building up to welcome however many thousand librarians to the city, and I'm sure you will enjoy the experience.

References

- International Coalition of Library Consortia (ICOLC) (2001) *Guidelines for Statistical Measures of Usage of Web-based Information Resources. Revised ed. December*. Available at <http://www.library.yale.edu/consortia/2001webstats.htm> (3 April 03)
- National Information Standards Organization (NISO) (2002). NISO Z39.7-2002. *Information Services and Use: Metrics and statistics for libraries and information providers – Data Dictionary [draft]. Review period: July 26, 2002-July 31, 2003*. Available at <http://www.niso.org/emetrics/> (3 April 2003).
- Shim, Wonsik “Jeff” et al (2001) *Measures and Statistics for Research Library Networked Services: Procedures and Issues. ARL E-Metrics Phase II Report*. Washington DC: Association of Research Libraries (ARL), October, p 15. Available at <http://www.arl.org/stats/newmeas/emetrics/phasetwo.pdf> (3 April 03).