Project COUNTER

Peter T. Shepherd

Project COUNTER

Peter T. Shepherd

Abstract

COUNTER is now the most widely adopted standard governing online usage statistics and has had a major positive impact on both vendors and librarians. The COUNTER Codes of Practice are refined and upgraded at regular intervals and this is necessarily a major focus for COUNTER. There is also another, growing area of COUNTER activity that is gaining in importance as the body of reliable, COUNTER-compliant usage data increases and as usage statistics feature more prominently in assessments of the value, status and impact of online publications. This area may be termed 'usage bibliometrics' and COUNTER is working actively with other organizations on the development of new, usage-based measures of value, status and impact. This article not only summarises the latest upgrades to the COUNTER Codes of Practice, but also provides examples of further improvements that are actively being investigated. In addition are described two research projects, in which COUNTER is involved, that are developing new, usage-based metrics.

Introduction

The last few years have seen not only great improvements in the reliability of usage statistics for librarians, but also the emergence of usage statistics as a central component of the librarian's management toolkit. But the challenge of ensuring that usage statistics continue to be relevant is considerable and ongoing. Constant attention is required on a number of fronts. Technology continues its relentless march and the way the usage of online publications is recorded and reported must take this into account; what was a meaningful metric five years ago may be less meaningful in the current technology environment. The main purpose of most librarians is not, we must remind ourselves, the collection and management of usage statistics; the time they can devote to this exercise is limited and the process should be made as efficient as possible. The current global economic downturn has, inevitably, increased the pressure on library budgets in both the public and private sectors and tough decisions will have to be made on collections in the next few years; usage statistics should help ensure that these decisions are well-informed and evidence-based.

Finally, the technical and business models for online publishing, especially of journals, are undergoing a significant shift and usage statistics must take this into account; online journals are more than simply collections of articles, while the growth of open access publishing means that journal articles are becoming more widely distributed with repositories as well as publishers hosting them.

COUNTER (1) was launched in 2002 with a mission to take a leading role in the development and implementation of online usage statistics, initially covering journals and databases, but subsequently expanding to cover online books and reference works also. COUNTER is now the most widely adopted standard governing online usage statistics and has had a major positive impact on both vendors and librarians. Vendors now have a practical standard governing the implementation of online usage statistics for their major product lines (journals, databases and books), while librarians now have access to reliable usage statistics for a significant proportion of the online content that they purchase. The COUNTER Codes of Practice, as well as the lists of COUNTER compliant vendors, are freely available on the COUNTER website[1]. But while the standards developed by COUNTER are set in stone, they have not and must not become petrified. Technology moves on, user behaviour changes and COUNTER must evolve with them. For these reasons the COUNTER Codes of Practice are refined and upgraded at regular intervals and this is necessarily a major focus for COUNTER. But there is another, growing area of COUNTER activity that is gaining in importance as the body of reliable, COUNTER-compliant usage data increases and as usage statistics feature more prominently in assessments of the value, status and impact of online publications. This area may be termed 'usage bibliometrics' and COUNTER is working actively with other organizations on the development of new, usage-based measures of value, status and impact.

Keeping the Codes of Practice up to date

Release 3 of the COUNTER Code of Practice for Journals and Databases was published in August 2008, the deadline for its implementation was September 2009 and over 100 vendors are now compliant with it. The main objectives of Release 3 of the Code of Practice are: first, to improve further the reliability of COUNTER usage reports by incorporating new protocols designed to mitigate the potentially inflationary effects on usage statistics of federated and automated search engines, internet robots, crawlers, etc.; second, to provide tools that will facilitate the consolidation,

management and analysis of COUNTER usage statistics; third, to improve the COUNTER usage reports for library consortia; fourth, to improve the reporting of the usage of journal archives. The main new features in Release 3 are listed below:

- The SUSHI (Standardized Usage Statistics Harvesting Initiative) protocol has been incorporated into the COUNTER Code of Practice. SUSHI has been developed by NISO (National Information Standards Organization) in co-operation with COUNTER and in 2007 became a NISO standard (Z39.93). Implementation of the XML-based SUSHI protocol by vendors will allow the automated retrieval of the COUNTER usage reports into local systems, making this process much less time consuming for the librarian or library consortium administrator.

- Usage Reports must be provided in XML, in addition to the existing prescribed formats (Excel, CSV, etc.). A link to the required SUSHI XML schema is provided below the Excel example of each usage report.

- Vendors that provide journal archives as a separate acquisition from the current journals must provide *either* **Journal Report 1a: Number of Successful Full-text Article Requests from an Archive by Month and Journal** (which was an optional additional usage report in Release 2) OR **Journal Report 5: Number of Successful Full-Text Article Requests by Year-of-Publication and Journal**

- In Database Report 1 and Database Report 3, search and session activity generated by federated search engines and automated search agents must be reported separately as illustrated in the example reports provided.

- New library consortium usage reports. The advent of the SUSHI protocol greatly facilitates the handling of large volumes of usage data, which is a particular advantage for consortial reporting. For this reason COUNTER has developed two new reports for library consortia that are specified only in XML format.

- A new protocol that requires federated and automated searches to be isolated from bona fide searches by genuine users, and reported separately in Database Report 1 and Database Report 3. The growing use of federated and automated searches has the potential to inflate enormously the

search and session counts in the database reports and this protocol is designed to mitigate such inflation.

- New protocols that require activity generated by internet robots and crawlers, as well as by LOCKSS and similar caches, to be excluded from the COUNTER reports

- A new optional additional report, **Journal/Book Report 1: Number of Successful Full-text Item Requests by Month and Title**, specified in XML only, will allow vendors that provide online journals and books on the same platform to report usage of both categories of product in a single COUNTER report.

Yet, despite these refinements the Code of Practice remains to some extent inadequate to the challenge of reporting database usage and we are actively investigating new approaches to the reporting of such usage. Also, new technologies and changing user behaviour threaten to reduce the usefulness of some of the established COUNTER reports and ways of dealing with these are also being investigated. It should be noted, however, that no new reports will be included in the COUNTER Codes of Practice until they have been fully evaluated by the community and approved by the COUNTER Executive Committee.

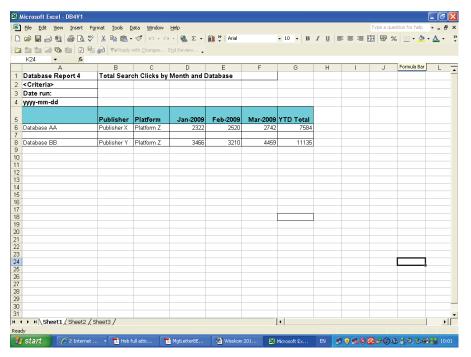### New approaches to measuring the usage of databases

COUNTER is investigating the development of a new database report that would supplement the existing COUNTER reports. We began by asking what the current Searches- and Sessions-based metrics tell us. Searches indicate the intensity of use of a particular database (how many queries are submitted) and Sessions the popularity of the database (how often users return to use it). Neither of these metrics gives us a true view of the *utility* of the given database--in other words, whether the searches carried out by the user have provided a useful return.

To help address the question of utility, COUNTER is investigating an additional metric, which JICWEBS (The UK Joint Industry Committee for Web Standards (2) calls 'Search Clicks' and is defined as 'A Click originating from a set of Search results'. This metric is an indicator of how many items offered by the result set (the Search) prompt user action to investigate further. Each such click is a request for a content item, most probably a full-text article - the "destination URL". If this report is imple-

mented the vendor would be required to identify all valid requests for destination URLs that have originated from a search results page. The resulting usage data would be subject to the usual COUNTER strictures about double-clicking, valid HTTP status and exclusion of robotic activity. Preparing the usage totals should be a simple matter of event processing in the same way as in the current COUNTER Journal Report 1.

Following a survey of librarians and vendors, COUNTER has developed Database Report 4: Total Search Clicks by Month and Database, which is provided in draft form below. This, perhaps with some modifications, will be published on the COUNTER website as an optional additional report. This approach allows vendors and librarians to assess how the new report works in practice, and to determine whether it provides them with valuable new insights into database usage. If we find that it is so, this report is may be upgraded to a *required* report in the next Release of the COUNTER Code of Practice.

Draft Database Report 4: Total Search Clicks by Month and Database

Note: Columns H (YTD Full-Text) and I (YTD Other) in the above example are designed to distinguish search clicks that yield a full-text article from other results (such as titles or abstracts).

While this is a useful distinction, it is not yet clear how many vendors can make it. Further feedback on this aspect will be obtained during the

period when this report is optional.

Dealing with changes in technology and user behaviour

While it is recognised that automatic and semi-automatic download tools, such as Quosa and PubGet, provide a very valuable service by facilitating access to full-text articles and other items of interest to the user, they may in some situations have an inflationary effect on the usage statistics reported by COUNTER-compliant vendors in the COUNTER usage reports. COUNTER's objective is to ensure the reporting of only genuinely user-driven usage and for this reason we think it is important that all other usage be identified and either reported separately, or, in some instances, excluded from the COUNTER usage reports. A protocol to cover these issues is currently being investigated.

Usage bibliometrics

There are currently two research projects in which COUNTER is involved that are investigating the development of new metrics derived from the COUNTER usage data. These are the Journal Usage Factor (JUF) project, which is assessing the feasibility of a usage-based equivalent of the citation-based Journal Impact Factor, and PIRUS2, which is developing a standard for recording and reporting usage at the individual article level.

Journal Usage Factor

ISI's journal Impact Factors, based on citation data, have become generally accepted as a valid measure of the quality of scholarly journals, and are widely used by publishers, authors, funding agencies and librarians as measures of journal quality(3). There are, nevertheless, misgivings about an over-reliance on Impact Factor alone in this respect and other, author-centred, citation-based measures, such as the Hirsch Index (4) are gaining support. The availability of the majority of significant scholarly journals online, combined with the availability of increasingly credible COUNTER-compliant online usage statistics, raises the possibility of a parallel usage-based measure of journal performance becoming a viable additional metric. Such a metric, which may be termed 'Usage Factor', could be based on the data contained in COUNTER Journal Report 1 (Number of Successful Full-text Article Requests by

Month and Journal) calculated as illustrated in Equation 1 below for an individual journal:

(1) Usage Factor = $\dfrac{\text{Total usage (COUNTER JR1 data for a specified period)}}{\text{Total number of articles published online (during a specified period)}}$

There is growing interest in the development of usage-based alternatives to citation-based measures of journal performance and this is reflected in the funding being made available for this work. Especially noteworthy in this respect is the work of Bollen and Van de Stempel (5).

Against this background the United Kingdom Serials Group (UKSG) thought it would be timely to sponsor a study to investigate the feasibility of journal Usage Factors.

The overall objective of this study is to determine whether the JUF concept is a meaningful one, whether it will be practical to implement and whether it will provide additional insights into the value and quality of online journals. The work has been carried out in two Stages: Stage 1 (Market Research) and Stage 2 (Modelling and Analysis).

Stage 1 of this project (6), a survey into the feasibility of developing and implementing a new metric–the Journal Usage Factor (JUF)--demonstrated not only that the concept is a meaningful one, but also that there is considerable support from the publisher, librarian and research communities for this new metric. UKSG, in association with the Research Information Network (RIN), and several publishers have provided funding to take this forward by undertaking a Stage 2 study to explore further the practical issues associated with the implementation of a Journal Usage Factor derived from COUNTER usage data.

The Stage 1 market research showed that the majority of publishers are supportive of the JUF concept, appear to be willing, in principle, to participate in the calculation and publication of Journal Usage Factors, and are prepared to see their journals ranked according to JUF. On the other hand, there is a diversity of opinion on the way in which JUF should be calculated, in particular on how to define the following terms: 'total usage', 'specified usage period', and 'total number of articles published online'. The Stage 2 modelling with real usage data will help refine the definitions for these terms.

Stage 1 also revealed that the great majority of authors in all fields of academic re-search would welcome a new, usage-based measure of the value of journals and there is not a significant difference between authors in different areas of academic research on the validity of journal Impact Factors as a measure of quality JUF, were it available, would be also be a highly ranked factor by librarians, not only in the evaluation of journals for potential purchase, but also in the evaluation of journals for retention or cancellation.

Stage 2 of the project will be to develop a programme of data modelling and analysis that uses real data from a number of publishers, with the aim of identifying potential candidate metrics for longer term, scaled-up testing. The work is being carried out by Frontline/John Cox Associates and the final report was published in August 2010. This report, together with further information on the JUF project, may be found on the UKSG website (7).

PIRUS: Publisher and Institutional Repository Usage Statistics

Until now the most granular level at which COUNTER requires reporting of usage is at the individual journal level. Demand for usage statistics at the article level from users has, until recently, been low. This, combined with the unwieldiness of usage reports in an Excel environment, has meant that COUNTER has given a lower priority to usage reports at the individual article level. A number of recent developments have, however, meant that it would now be appropriate to give a higher priority to investigating the development of a COUNTER standard for the recording, reporting and consolidation of usage statistics at the individual article level. Most important among these developments are:

- The growth in the number of journal articles hosted by institutional and other repositories, for which no widely accepted standards for usage statistics have been developed
- The emergence of online usage as an alternative, accepted measure of article and journal value and usage-based metrics being considered as a tool to help assess the impact and value of publications
- Authors' and funding agencies' increasing interest in a reliable, global over-view of usage of individual articles

- The fact that some publishers, notably PLoS, have already implemented the reporting of usage at the individual article level and are actively seeking a global standard on which they can base such reports
- Implementation by COUNTER of XML-based usage reports that make more granular reporting of usage a practical proposition
- Implementation by COUNTER of the SUSHI (2) protocol, which facilitates the automated consolidation of large volumes of usage data from different sources.

The original PIRUS project, funded by JISC, (the United Kingdom Joint Information Systems Committee) and completed in January 2009 (8) demonstrated that it is technically feasible to create, record and consolidate usage statistics for individual articles using data from repositories and publishers, despite the diversity of organizational and technical environments in which they operate.

The four main outcomes of the project were:
- A proof-of-concept COUNTER-compliant XML prototype for an individual article usage report, Article Report 1: Number of Successful Full-text Article Downloads that can be used by both repositories and publishers. In principle this report could be provided for individual authors and for institutions. In practice, the individual author reports are much easier to generate and are a realistic short-term objective, while the reports for institutions and other entities, such as funding agencies, will be more complex and should be regarded as a longer term objective.
- A tracker code, to be implemented by repositories, that sends a message either to an external party that is responsible for creating and consolidating the usage statistics and for forwarding them to the relevant publisher for consolidation or to the local repository server
- A range of Scenarios for the creation, recording and consolidation of individual article usage statistics that will cover the majority of current repository installations. Each repository may select the scenario that corresponds to its local technology and implementation.

- A specification for the criteria that would have to be met by a central facility that would create the usage statistics where required (for some categories of repository) and collect and consolidate the usage statistics for others.

If these outcomes are to be translated into a new, implementable COUNTER standard and protocol, further research and development is required, specifically in the following areas:

- <u>Technical:</u> further tests, with a wider range of repositories and a larger volume of data, will be required to ensure that the proposed protocols and tracker codes are scalable/extensible and work in the major repository environments.
- <u>Organizational:</u> the nature and mission of the central clearing house/houses proposed by PIRUS have to be developed, and candidate organizations identified and tested
- <u>Economic:</u> assess the costs for repositories and publishers of generating the required usage reports, as well as the costs of any central clearing house/houses; investigate how these costs could be allocated between stake holders
- <u>Political:</u> the broad support of all the major stakeholder groups (repositories, publishers, authors) will be required. Intellectual property, privacy and financial issues will have to be addressed.

PIRUS2

PIRUS 2, also funded by JISC, was launched in October 2009 as a co-operative project involving publishers and repositories. Building on the work of PIRUS 1, it will develop a set of standards, protocols and processes to enable publishers, repositories and other organizations to generate and share authoritative, trustworthy usage statistics for the individual articles and possibly other items that they host.

PIRUS 2 has set the following specific objectives:

- The development of a suite of free, open source programmes to support the generation and sharing of COUNTER-compliant usage data and statistics that will cover individual items in publisher, aggregator, institutional and subject repositories
- The development of a prototype article-level Publisher/Repository usage statistics service
- Defining a core set of standard usage statistics reports that publishers and repositories could produce for internal and external consumption
- Assessing the costs for publishers and repositories of generating the required usage reports, as well as the costs of any central clearing house/houses; investigate how these costs could be allocated between stakeholders

To achieve these objectives, the project is organized into six workpackages, described in detail in the project plan, which is available on the PIRUS 2 project website (9)

The work of PIRUS2 will ensure that usage data are available for journal articles wherever held (publisher sites, repositories, aggregators), whilst going further than web analytics software and more able to meaningfully address the consistency of the usage data and the resultant quality of the reports. Repositories will benefit from a technical point of view as PIRUS2 will provide them with access to new functionality to produce standardised usage reports from their data. Digital repositories systems will be more integral to research and closely aligned to research workflows and requirements, as the project addresses production of authoritative usage data. The authoritative status of PIRUS 2 usage statistics will serve to enhance trust across repositories; furthermore, the data will provide a firm evidence base for repositories to take firm steps to defining clear policies to support their goals.

PIRUS 2 is lead by Mimas (The University of Manchester) and Cranfield University; the other primary partners are COUNTER, Oxford University Press and CrossRef. Oxford University, Southampton University and other Institutional Repositories will also participate in PIRUS 2, together with publishers, repositories and related projects in the UK, USA and elsewhere. Work on PIRUS 2 commenced in October 2009 and the final report will be published in December 2010.

Further information and updates on the project may be found on the PIRUS2 website (see reference[9]).

COUNTER: an ongoing challenge

In the eight years since it was launched, COUNTER has become central to the planning of online publishers and librarians, initially in the academic/research world, but increasingly in the business publishing world. The COUNTER reports are now an essential tool in demonstrating the value of individual publications and entire online collections. More recently, it has become apparent that there are potentially valuable new metrics that could be derived from the growing body of COUNTER usage statistics. COUNTER's work is ongoing, and constant attention is required to maintain the structure.

References

1. Project COUNTER: www.projectCounter.org

2. JICWEBS: http://www.jicwebs.org/jicwebs_results.php

3. For a discussion of issues related to Impact Factor, see Cross, J, Impact Factors—The Basics. In: *Electronic Resources Management Handbook*: http://tinyurl.com/2p935y

4. Hirsch, J.E., PNAS, 102 (2005), pp16569-16572. An index to quantify an individual's scientific research output http://www.pnas.org/cgi/content/abstract/102/46/16569

5. Bollen, J & Van de Sompel, H: http://arxiv.org/abs/cs.DL/0610154

6. Journal Usage Factor Stage 1 Final Report: http://www.uksg.org/usagefactors/final

7. Journal Usage Factor progress reports: http://www.uksg.org/usagefactors/progress

8. PIRUS1 Final Report http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus_finalreport.pdf

9. PIRUS2 Progress Reports: http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/