Usage statistics and online behaviour

ANGELA CONYERS
Senior Research Fellow, UCE Birmingham

Introduction

This chapter sets out to explore the variety of statistics now available to measure use of electronic resources, and some of the issues connected with these statistics. It will look at how reliable usage statistics can form the basis of a range of analyses at both local and national level and how usage statistics combined with other data can present a picture of how online services are being used. The chapter is based on research within the UK higher education community, but the findings are applicable to any library or group of libraries that wishes to explore more fully their use of online resources.

Evidence base is a research and evaluation unit within Library Services at UCE Birmingham and this discussion draws mainly on the findings of two research projects currently being undertaken there: the e-measures project, funded by the Higher Education Funding Council for England (HEFCE), and the NESLi2 study and analysis of usage statistics, funded by the Joint Information Systems Committee (JISC). A description of these two projects is given below.

E-measures project

The e-measures project\(^1\) is funded by HEFCE under the Good Management Practice programme and aims to assist libraries with the use of statistics for decision-making and user support. Evidence base is also working with the Society of College, National and University Libraries (SCONUL) to devise a set of performance indicators for electronic resources for inclusion in the SCONUL Annual Statistics.\(^2\) Twenty-five higher education libraries took part in a pilot project during 2003–4, setting up systems for data collection and collecting data over a four-month period (February–May 2004) for a range of 20 e-measures covering the holdings, usage and costs of electronic resources. In addition to the amount of data collected from such a large sample of libraries, the project highlighted the variety of methods in use for data collection and the issues that libraries faced in attempting to obtain reliable and consistent data.\(^3\)

As a result of the work on this pilot project, a set of questions on electronic resources is now included in the SCONUL statistical return which all UK higher education libraries are asked to complete. These questions include two relating specifically to usage, covering full-text serials and e-books. The SCONUL annual statistics have long provided a useful guide to overall trends within academic libraries, and a source of benchmarking for groups of libraries. The new e-measures questions will, over time, produce a...
picture of the extent of the shift to electronic delivery in the UK higher education sector and enable performance ratios to be established. The questions developed for SCONUL also took account of the work done by the Association of Research Libraries (ARL) in the United States, and of international standards, thus facilitating international comparisons.

**NESLi2 study and analysis of usage statistics**

The NESLi2 study aimed to provide the JISC and its Journals Working Group (JWG) with accurate and up-to-date data on national use of journals available through the NESLi2 initiative. This is the UK national scheme for the licensing of e-journals on behalf of the higher and further education and research communities. Evidence base undertook this work in association with staff from Cranfield University and from Florida State University and with the Director of COUNTER. The study took usage data from four major NESLi2 publishers for 17 higher education libraries over an 18-month period (January 2003–June 2004) and analysed these in relation to cost, price, subject category and usage range. A framework for analysis was devised that could be further developed with other libraries and publishers. The study provided more evidence of the difficulties inherent in the interpretation and analysis of usage statistics and the amount of time needed to achieve meaningful results. Detailed recommendations arising from this study are currently under discussion with the JWG and the study itself is being extended to cover the whole of 2004.

In order to respect confidentiality agreements with publishers and libraries, the results of this study are not publicly available. A summary report can be found on the NESLi2 web site.

**Why collect usage statistics?**

Usage statistics provide an excellent example of information overload. Prior to the advent of e-journals, usage of journal collections was a virtually unknown factor, in spite of librarians’ attempts over the years to devise various forms of survey. Now there appears to be no limit to the amount of usage data publishers can provide, down to the individual page level. Data provided by publishers is supplemented by that from aggregators and gateways, Eduserv Athens, library system suppliers, and a range of other services designed to make the management of electronic resources simpler. As it is easy for publishers and suppliers to produce statistics, the library gains access to more and more raw data with endless opportunities for analysis. Seeing the wood from the trees provides an appropriate analogy for the resulting volume of print-outs that could be produced.

In this environment, it is important first of all to determine the purpose for which usage statistics are needed as this will help decide what statistics to use and how they are to be analysed and presented. Some possible reasons are given below, with examples of questions that can be asked.

*Library promotion and publicity*
- How many requests for a particular group of resources have there been over a specified period?
- What is the percentage increase in usage over time?

*User support*
- Are there particular resources that are not being used as much as expected?
- Should more attention be given to promotion?
- Which are the most popular routes for access? (Directly from publisher, A–Z, library pages, through library catalogue, etc?)

*Making renewal decisions*
- How much has the resource been used?
- What is the average cost per request?
How many titles in the deal are receiving nil or low use?
How do average costs and total costs compare with other deals?

**Budgeting**
- How do usage statistics demonstrate value for money and justify the library’s expenditure?
- What is the balance of spending and usage in different subject areas?

**Benchmarking**
- How does the library compare in terms of number of requests and costs per request with other libraries of similar type or size?

**Contribution to national overview**
- What trends are emerging nationally on use of electronic resources?
- How do these relate to statistics on traditional library services such as number of book issues, number of inter-library loans, etc?

While usage statistics alone cannot provide a complete answer to all the questions posed above, they can certainly contribute to any analysis and inform decisions on how results can be presented.

**Defining the type of resource**

Before the e-measures project, SCONUL had used an all-encompassing ‘views/downloads’ question in its statistical return to get a first feel of what was happening in online services. This provided a useful starting point and, given the lack of reliable statistics at the time, there was little else to be done. The problem was that libraries interpreted the question in different ways (if they answered it at all) and produced vastly different results, so the answers were very far off the sort of benchmarking figures that the SCONUL statistics give in such traditional library areas as book issues or gate counts.

For library publicity purposes, large global figures may be used to good effect, but usage statistics are rarely used on their own in this way. When further analysis is required, such as ‘cost per request’ or ‘average use per title’, then usage statistics must be tied in with relevant cost and holdings information, so that appropriate ratios can be derived. For this reason, usage statistics for different types of resource need to be viewed separately.

The e-measures project identified six different types of electronic service provided by the library and set up usage measures for each:

- serials
- databases
- e-books
- digital documents
- virtual visits: number of hits on the library web site
- enquiries received electronically.

There was much discussion in the project about what constituted a collection of serials or e-books (where titles would be counted individually) and what was a database (whose individual titles would not be counted separately). A definitions table9 has been produced using guidance from international standards wherever possible. The aim has been to produce a consistent approach for the SCONUL statistical return, which would ensure that as far as possible all libraries were counting the same thing, and thus avoid the wide variations noted in the previous ‘views/downloads’ question.
Identifying the source of usage statistics

Publisher or vendor
For serials, databases and e-books, the primary source of usage statistics is the publisher. These statistics are generally available on a monthly basis from the publisher or vendor web site using password-protected access, though other methods may also be used.

Gateways, hosts and aggregators
If the library gives access to e-serials via a gateway service such as SwetsWise, a hosting service such as Ingenta or an aggregator such as ProQuest, then separate usage statistics will generally be provided. These will give a record of the volume of usage through the particular service, but will not include any requests made directly to the publisher. Some publishers include these requests within their own usage statistics, others do not. A list provided by Swets is available on the evidence base web site indicating for which publisher SwetsWise statistics are to be added.

Eduserv Athens
The Eduserv Athens authentication service is another important source of statistics, particularly for databases where no other source may be available. Libraries in the e-measures project reported on this use of Athens statistics but remarked on differences when comparing results with those from publishers, which indicated that the counts were not working in the same way. Some libraries are using Athens statistics to show how much usage is coming from a particular source, e.g. on or off the campus, or from a particular group of users. This provides good management information of a type not available from publisher usage statistics. Its use for this purpose will depend on the way user groups are identified through passwords and on how access through Athens authentication is organized for on-campus and off-campus users.

Web-logging software
For gathering information on usage of digital documents produced in house, on virtual visits or hits on the library web site, and on enquiries received electronically, it is necessary to have appropriate web-logging software, such as WebTrends, installed on the relevant servers. A few libraries in the e-measures project had succeeded in setting up systems which could measure on- and off-campus use of the library web site, but results reported were not compatible between libraries, indicating that this was a measure that could only be used in house to show usage over time. Even then, it could prove problematic as there may be elements of double-counting. Certain libraries had developed in-house systems for measuring use of digital documents. In future, integration of such services into library systems or into virtual learning environments (VLEs) is likely to make more standardized statistics easier to obtain.

Library management systems
Increasingly, library systems suppliers are providing means of linking directly to electronic resources. Usage statistics from such sources will provide useful guides to the amount of traffic going through the particular route, but will not pick up on requests for e-journals or other electronic resources made directly to the publisher or vendor web site.

Publishers’ usage statistics (supplemented where necessary by gateway or aggregator statistics as noted above) are therefore the primary source of usage statistics, although other sources also noted above may provide valuable supporting information or provide statistics where no other routes are available.
Finding consistent and reliable usage statistics

At the start of the e-measures project, pilot members were unanimous in describing the difficulties they had with inconsistent, unreliable usage statistics. One of their main aims was to see an improvement in the quality of data so that performance and usage could be compared with a greater degree of confidence. The work of COUNTER (Counting Online Usage of Networked Electronic Resources) was therefore seen as an encouraging sign of improvements to come.

COUNTER

The COUNTER initiative provides an excellent example of an international co-operative project involving vendors, librarians and intermediaries. The project looked at the range of publisher usage statistics available and aimed to introduce some standard definitions and reports that could give libraries confidence that they were comparing like with like when looking at usage statistics from different suppliers or even looking at trends within the same deal over time. COUNTER was launched in March 2002 and issued its first Code of Practice for journals and databases in December 2002.

Usage measures for the e-measures project were chosen wherever possible to coincide with the relevant COUNTER report. The project’s four-month data collection phase (February–May 2004) thus provided an opportunity to test the success of COUNTER among a group of libraries and to see how many publishers were now providing ‘COUNTER-compliant’ reports.

COUNTER reports for journals

For journals, the COUNTER Code of Practice has four reports:

- JR1 Number of successful full-text article requests by month and journal
- JR2 Turnaways by month and journal
- JR3 Number of successful item requests and turnaways by month, journal and page type
- JR4 Total searches run by month and service.

The latter two are not mandatory for the Code of Practice.

In the latest list of COUNTER-compliant vendors (dated May 2005) 38 publishers and intermediaries are listed as COUNTER compliant in respect of the JR1 report. This compares to 15 at the start of the e-measures project in October 2003, and demonstrates the success of COUNTER in persuading most of the larger providers in the e-journal market to sign up for compliance. As a further encouragement for publishers to be COUNTER compliant, this is now a requirement for deals negotiated for UK academic libraries through the NESLi2 initiative. The process has helped librarians to have greater confidence in serials usage statistics as being comparable across publishers and over time.

In the e-measures project, libraries were asked to identify what proportion of usage statistics for serials were COUNTER compliant. Seventeen of the 25 libraries provided this breakdown and gave an average estimate of 79%. This gave some indication of the extent of COUNTER compliance, which encouraged the e-measures team to recommend to SCONUL that the successful full-text article request (COUNTER JR1) be one of the two usage statistics that libraries are asked to collect at a national level for benchmarking and trend analysis over time. Similarly, the NESLi2 study used the number of successful full-text article requests as the basic unit of measure in its analysis of usage statistics.

COUNTER reports for databases

There are three COUNTER reports for databases:

- DB1 Total searches and sessions by month and database
- DB2 Turnaways by month and database
- DB3 Total searches and sessions by month and service.
In contrast to the 38 publishers providing JR1 reports, there were 12 suppliers listed in May 2005 as providing DB1 reports and 14 providing DB3 reports. This compares to six and eight respectively in October 2003. While this marks an improvement, there are still some notable omissions, for example in the law area. One library in the e-measures project estimated that only around 60% of their databases had any kind of statistics available.

In the e-measures reports on database sessions and searches, there was a large variation in the number shown as being COUNTER compliant, with an average estimate of 32% from the 21 libraries responding to this question.

**COUNTER reports for e-books**

The COUNTER Code of Practice for books and reference works has been available for consultation in draft form since January 2005 and the final version will be formally launched early in January 2006. This new Code of Practice proposes five reports for books and reference works:

- BR1 Number of successful title requests by month and title
- BR2 Number of successful section requests by month and title (a section is defined as a chapter or entry)
- BR3 Turnaways by month and title
- BR4 Total searches and sessions by month and title
- BR5 Total searches and sessions by month and service.

The new COUNTER code is to be welcomed in trying to introduce the same consistency into usage statistics for e-books as for e-journals and databases.

**Issues in the collection of usage statistics**

**Serials**

While the successful full-text article request (COUNTER JR1) can now be regarded as a standard measure, both the e-measures and the NESLi2 studies identified ongoing problems for libraries in collecting and analysing serials statistics:

- **Time required**
  The process of getting passwords for individual access to publisher web sites is time-consuming. Some publishers require e-mail requests. A guide to sources of statistics produced for the e-measures project is available from the SCONUL web site, but this needs continual updating as services change. It is hoped that eventually there will be a central source for all publisher usage statistics.

- **Incomplete statistics**
  There are still some vendors who do not provide regular statistics; collecting usage figures for individual ‘free with print’ titles was seen by many libraries as too time-consuming.

- **Lateness of data**
  Data was not always presented on time; this raises the question of how often and on what basis usage statistics available monthly from publishers should be collected and analysed.

- **Gateway statistics**
  The COUNTER Code of Practice identifies and defines ‘hosts’ such as Ingenta, ‘gateways’ such as SwetsWise and ‘aggregators’ such as ProQuest as alternative sources of usage statistics. These services may also be providing COUNTER-compliant JR1 reports and it was not always clear whether these should be added.

- **Inconsistency of data**
  There were problems in matching titles over a certain time-period because lists change and are updated where serials move publishers or new titles are added.
**Databases**

In producing figures for database searches and sessions, e-measures libraries reported the following issues:

- lack of usage data from some major providers
- inconsistency/unreliability of data
- relationship between data provided by Athens and that from suppliers.

Because of this uncertainty and the more limited success with COUNTER compliance noted above, it was decided not to include database searches and sessions in the SCONUL return. This does not mean of course that libraries should cease to collect this data. In fact it is hoped that they will aim to do so and thus put pressure on more database providers to provide COUNTER-compliant reports.

**E-books**

E-measures libraries were asked to record monthly totals for successful accesses to e-books. Usage statistics for e-book collections were obtained from the vendors, and usage of individual e-books from web-logging software or library management systems.

During the four-month data collection phase, a number of libraries were adding to their e-book collections. Issues reported were:

- incomplete data, especially when no web-logging software was in place
- use of different terminology by different vendors
- manipulation of data was sometimes needed to get results.

The new COUNTER Code of Practice for books and reference works is therefore to be welcomed. When the e-measures project asked heads of library services what measures they would like to see included in the SCONUL statistical return, there was strong support for the inclusion of a measure of use for e-books. In spite of the problems noted above, it was therefore decided to include a question on ‘number of successful accesses to electronic books’ in the SCONUL statistical return.

This is a growing area within academic libraries and being ‘in on the ground floor’ will be an advantage in benchmarking and showing trends if and when use of e-books ever reaches comparable figures to physical book issues.

**Analysis of usage statistics**

Through the e-measures project, therefore, it is possible to demonstrate many of the difficulties inherent in the collection of usage statistics and to point to some of the ways forward. Collecting usage statistics, whether nationally for SCONUL or locally within the library, is certainly not currently as straightforward as counting book issues. On the other hand, the amount of data now available can provide a much clearer picture of online behaviour. The NESLi2 study seeks to show the ways in which usage statistics can be analysed to show how the ‘big deals’ negotiated for the higher education community are actually being used and how value for money can be demonstrated.

As described above, the NESLi2 study took as its starting point the number of successful full-text article requests (COUNTER JR1) for selected publishers and libraries and used these as the basic unit of measure. Usage statistics on their own can illustrate trends over time and patterns of use through the academic year, but for further analysis it is necessary to apply other variables in order to establish performance ratios. The number of requests was looked at in relation to the following variables:

- **Usage range**
  The number of successful requests for each title over a given period was sorted by usage range, with categories for nil, low, medium, high and very high usage. This made it possible to show the percentage of total requests falling into each usage range and give a measure of comparison across libraries and across publisher deals.
It is recognized that a certain number of titles within a deal receive very high use. This method allowed those titles to be identified and also showed what percentage of total requests were coming from this relatively small percentage of titles.

**Price range**
Using the publisher list prices, titles within a deal were divided into low, medium, high and very high price ranges. The same price ranges were used across all publisher deals studied, so that the percentage of high or very high price titles in each could be compared. It was then possible to look for a correlation between journal price and journal use, to see if high cost titles were used more and low cost titles used less.

**Subject category**
For this study, all titles were identified as either STM (science, technology and medicine) or HSS (humanities and social sciences), using categories given by the publisher or devised by the study team. Usage could then be analysed by subject category to see which category of title was being used most heavily and whether differences were consistent across libraries.

**Subscribed titles**
Information was obtained from libraries on the titles within a deal to which they held subscriptions. Since pricing of a ‘big deal’ depends heavily on the number of subscriptions previously held, this was an important variable in looking at usage. Were the subscribed titles being heavily used and their choice therefore justified? On the other hand, was there evidence of use of previously ‘unsubscribed’ titles to show how much these additional titles included in the deal were being used?

In addition to these methods of analysing usage statistics, further information was sought either directly from the libraries taking part or from the SCONUL or Higher Education Statistics Agency (HESA) statistics. This included:

**FTE users**
The study used a figure which combined full-time equivalent (FTE) student numbers and FTE academic staff numbers to produce a series of ratios relating to number of requests, number of titles, costs, etc.

**Contextual information**
The study used JISC banding as its primary division. The JISC groups higher education institutions into ten bands (A–J) and this banding generally forms the basis of subscription agreements with publishers. JISC banding was supplemented by type of institution (old or new university, HE College) and number of undergraduate and postgraduate students on STM- or HSS-related courses. The purpose of this was to look for any common characteristics which explained emerging usage patterns or, conversely, to see if there was any correlation between usage patterns and type of institution.

**Costs**
Of all derived ratios, those showing value for money are likely to be of the greatest interest. To establish ‘average cost per request’ the study team obtained information from each library on the total cost of the deal. This often meant collecting data from different sections of the library in respect of printed subscriptions obtained through an agent and e-access fees paid directly to the publisher.

The study team used the total cost of subscribed titles and the additional costs of e-access to all other titles included in the deal to break down the ‘average cost per request’ into ‘average cost per subscribed request’ and ‘average cost per unsubscribed request’. The team also developed the idea of yield per £. Comparing the list prices of ‘unsubscribed’ titles with the additional cost of e-access shows the amount that is gained for each £ of e-access expenditure.

Using these methods, it was possible to arrive at a set of average costs that could then be viewed across deals and across libraries and to show the link between usage and costs. These costs could then be compared with the costs of other methods of document delivery such as inter-library loan or pay-per-view to establish value for money.

**Nil use titles and availability**
It is a common complaint that the ‘big deal’ forces libraries to buy titles that they do not need and they do not use. A straight reading of usage statistics suggests that it is frequently the case that a large percentage
of titles are not used. However, a more detailed analysis of these ‘nil use’ titles may show that a significant proportion of them are not actually available in the deal because they have not yet been added or have recently been removed. In order to arrive at an accurate figure for nil use, those titles included in the list of usage statistics but not included in the deal need to be removed from the analysis. Further examination of remaining titles can then give a truer picture of non-use in relation to titles by price band and will indicate whether there is a real cause for concern.

This detailed analysis of usage statistics for the NESLi2 study was supplemented by conducting case-studies in selected libraries, seeking comments both on the reports produced for the case-study institution and on the deals themselves.

The methodology developed enabled a broad picture of online usage to be presented for the NESLi2 study and the framework devised could readily be adapted for other libraries and publisher deals, enabling individual libraries to analyse their own usage patterns, or groups of libraries to compare results against agreed ratios. A similar approach is also underway in a research project at the University of Newcastle Library where COUNTER-compliant usage data is being used to benchmark electronic resources.

Contextualizing the data

The type of analysis outlined above enables as much quantitative information as possible to be gleaned from any set of usage data. It does not, however, give any qualitative information in respect of the actual use that students and staff are making of the online services in which the library invests so heavily. If users have downloaded an article, how useful have they found it? In one respect this is no different from knowing when they borrow a book or photocopy a journal article whether they have actually read it or how useful it was. This should not, however, deter libraries from getting behind usage statistics to see what is going on, so that they can understand more about online behaviour. This will help inform strategies for information literacy training, technical support and promotion of services. Decisions on moving to an e-only environment or making print journals less accessible may provide opportunities for exploring the effect on users by focus groups, web surveys or other qualitative research methods. The support from UK higher education libraries to take part in the annual LibQUAL+ surveys demonstrates the importance attached to obtaining qualitative data on library use and possibly in future to benchmark it.

The Impact project, a joint initiative from SCONUL and LIRG (Library and Information Research Group), found measuring the impact of electronic resources was the topic of greatest importance among participants. Descriptions of the methodologies employed by participating libraries offer ideas for qualitative impact studies that other libraries may wish to follow.

The evaluated project and toolkit also developed by evidence base includes tools for qualitative analysis of e-library services and in this way complements the more quantitative approach taken by the e-measures project.

Conclusion

Although many issues remain, the emergence of more reliable usage statistics through the work of COUNTER now makes it possible for libraries to take a more systematic approach to their analysis and to build up a clearer picture of online behaviour from this statistical base. This will help individual libraries in their decision-making and user support, while the sharing of data through SCONUL or other library groupings will assist with benchmarking and the charting of trends in the use of electronic resources.
Web sites of organizations included

Evidence base: http://www.ebase.uce.ac.uk/
Higher Education Funding Council for England (HEFCE): http://www.hefce.ac.uk/
NESLi2: http://www.nesli2.ac.uk/
Joint Information Systems Committee (JISC): http://www.jisc.ac.uk/
Society of College, National and University Libraries (SCONUL): http://www.sconul.ac.uk/
JISC Journals Working Group (JWG): http://www.jisc.ac.uk/index.cfm?name=wg_journals_home
Eduserv Athens for Education: http://www.athens.ac.uk/
Ingenta: http://www.ingenta.com/
ProQuest Information and Learning: http://www.proquest.com/
WebTrends: http://www.webtrends.com/
Counting Online Usage of NeTworked Electronic Resources (COUNTER): http://www.projectcounter.org/
Higher Education Statistics Agency (HESA): http://www.hesa.ac.uk/
University of Newcastle Library: http://www.ncl.ac.uk/library/
CILIP (Chartered Institute of Library and Information Professionals)
Library and Information Research Group (LIRG): http://www.lirg.org.uk/

References

1. E-measures project: http://www.ebase.uce.ac.uk/emeasures
5. International Coalition of Library Consortia (ICOLC), Guidelines for statistical measures of usage of web-based information services: http://www.library.yale.edu/consortia/2001webstats.htm
6. NESLi2 study and analysis of usage statistics: http://ebase.uce.ac.uk/projects/NESLi2.htm
11. Project COUNTER:  
   http://www.projectcounter.org
12. COUNTER-compliant vendors:  
   http://www.projectcounter.org/articles.html
13. COUNTER draft Code of Practice on e-books and reference works:  
   http://www.projectcounter.org/cop_books_ref.html
14. E-measures sources of statistics:  
   http://sconul.ac.uk/activities/performance/papers/emeasuresources.doc
18. Evalued project and toolkit:  
   http://www.evalued.uce.ac.uk

© Angela Conyers

Dr Angela Conyers  
E-mail: Angela.Conyers@uce.ac.uk  
Correspondence address:  
Evidence base  
Research & evaluation  
UCE Library Services  
UCE Birmingham  
84 Aldridge Rd  
Perry Barr  
Birmingham B42 2SU, UK

Biographical note

Dr Angela Conyers is Senior Research Fellow at evidence base, UCE Birmingham. Before moving into research two years ago, she was for 13 years Director of Library Services at Canterbury Christ Church University College. She has worked recently on the e-measures project funded by HEFCE (Higher Education Funding Council for England) and on the NESLi2 study and analysis of usage statistics for the JISC (Joint Information Systems Committee). Both projects have involved working on usage statistics for electronic services with a large representative sample of UK higher education libraries.