
What do publisher usage statistics tell us?

The Analysing Publisher Deal project from Evidence Base

Dr Angela Conyers
Senior Research Fellow,
Evidence Base Research & Evaluation,
Library and Learning Resources,
University of Central England in
Birmingham
E-mail: angela.conyers@uce.ac.uk

INTRODUCTION

Over the past year, fourteen higher education libraries have been taking part in a self-funded project run by Evidence Base at UCE Birmingham with the support of SCONUL's Working Group on Performance Improvement (WGPI). The aim of the project is to assist libraries in the collection and analysis of e-journal usage statistics available through NESLi2 (the UK national e-journals initiative for further and higher education) or similar publisher deals.

Evidence Base Research and Evaluation, a research unit within UCE Birmingham Library and Learning Resources, had worked previously with the WGPI on the e-measures project¹ which had helped in deciding which new questions on e-resources should be added to the SCONUL questionnaire. It has also worked with the Joint Information Systems Committee (JISC) on a study and analysis of usage statistics for a selection of NESLi2 deals.² Both projects involved working with a wide range of higher education libraries. Having built up a good deal of experience of the issues involved and developed a methodology for analysing usage statistics, we were keen to see how we could provide libraries with some practical help and guidance in analysing and presenting their own data. We also wanted to give opportunities for sharing experience and benchmarking.

Anyone responsible for the collection of e-journal usage statistics will be aware of the pitfalls it can present. Without the COUNTER code of practice,³ a project such as this would not have been possible. One essential stipulation for the selection of publishers to be analysed was that they should provide COUNTER-compliant statistics so that we could be sure of a level of consistency across the deals analysed. While recognising the importance of an accurate basis for the presentation of statistics, our aim was to point out the issues but then move on to examine the way usage data could be used by management for collection development, budgeting and demonstrating value for money.

SETTING UP THE PROJECT

Fifteen libraries were initially recruited to the project, to start in early 2006 as a result of a call on lis-SCONUL. The project was designed to be self-funding, with libraries paying a small fee to take part. By the end of the project, those taking part would have:

- reports on their usage of up to six publisher deals over a two-year period (2004–5)
- ideas on how usage statistics could be analysed and presented with their own institutions in future
- the opportunity to share methods and ideas with other libraries and to benchmark results.

At the start of the project, participants were invited to submit a list of their publisher deals in priority order, the only stipulation being that the publisher must be COUNTER-compliant so that statistics were comparable. (COUNTER is Counting Online Usage of Networked Electronic Resources.) From this, a list of six publishers was drawn up. Benefits that participants hoped to gain from the project included:

- increased knowledge of what outputs can be used for
- assessment of the benefits of publisher deals
- analysis of usage against cost, value for money
- making evidence-based decisions
- the informing of purchasing decisions
- learning from others/contributing their own experience
- sharing best practice – and frustrations
- benchmarking results with comparable institutions.

It was clear that, in addition to the above benefits of participation, some libraries saw the project as an opportunity to develop staff and build staff skills and capacity.

Initial workshops were held in London and Birmingham early in 2006, with participants being introduced to the methodology adopted for the project and working through some examples using Excel and Access. The experience of working with these groups helped in the design of the manuals to be used for the project, as it emphasised the need for clear instructions that could be followed whatever the level of initial knowledge. E-mail and phone support was available to all participants throughout the project.

TEMPLATES AND MANUALS

All analysis was based on the COUNTER JR1 (Journal Report 1) report 'Number of successful full-text article requests by month and journal'. Further contextual information was provided by the library or by the SCONUL annual library statistics relating to the total cost of the deal, the size of the library serials budget and the number of FTE (Full Time Equivalent) students and teaching staff. For each publisher selected for the study, participants were given access to the following tools provided by Evidence Base on a secure shared-file area:

1. An Excel spreadsheet template: this listed all titles included in the deal for 2004 and 2005 and was set up so that the library's own usage data could be added in and analysed by usage range (nil, low, medium and high use), by broad subject area (humanities and social sciences (HSS) or science, technology and medicine (STM)) and by price band (unpriced, low, medium or high cost). Libraries also added details of their subscribed titles.
2. A Word report template: once the statistics had been set up in the spreadsheet and a set of tables produced from them, these tables could be transferred to a Word report which also gave suggestions of key points to look out for in presenting the analysis. Tables produced included:
 - monthly usage over the two-year period, showing trends and percentage changes
 - average, median and maximum number of requests per title

- costs per request, including costs of subscribed and unsubscribed titles, yield benefit of subscribing to the deal
- number of titles and requests in each usage range, analysed by subscribed and unsubscribed titles, price band and broad subject category
- details of titles with high use, nil or low use.

This template formed the basis of the report on two years' usage which was the final product of the analysis and which could be used both internally as a report to management and externally for benchmarking results with other members of the project.

3. Manuals: a set of manuals was produced, with step-by-step instructions on how to enter usage data, how to compile the tables and how to complete the report.

WHY HAVE A TEMPLATE?

The most straightforward approach to usage statistics is just to download the titles as they appear on the publisher's website, identify high, low and nil use titles and compare total requests with total costs to get the basic 'cost per request' which acts as a common yardstick of use.

While some usage measures can be derived in this way, a lot of opportunities for further analysis will be missed:

- Grouping titles by usage range: once the usage statistics are entered in the template, a formula is used to add up the monthly columns and group annual totals for each title into a series of ranges marking nil and low use (under 10 successful requests per title), medium (under 100 requests) and high use (100 requests and over). Further subdivisions can be used, for example to track very high-use titles (over 1,000 requests). This approach has the advantage of making it easy to see at a glance what percentage of titles have high, medium or low use. As described below, it can also form the basis of further analysis by price band or subject category.
- Matching prices to titles: for the spreadsheet template, publishers' price lists were downloaded and matched with titles in the deal. Prices were then sorted by price range, covering low price (under £200), medium (£200–399), high (£400–999) and very high

(£1,000 and over). The value of a deal can then be assessed in terms of the number of high-value titles it contains and a correlation can be sought between usage and price range. Are the high-priced titles the most used? Are titles with nil and low use mainly those which are low-priced?

- Identifying nil-use titles: a cursory glance at a set of publisher usage statistics will generally show a high proportion of titles with nil use recorded against them. This may lead to the view 'This deal is no good for us, half the titles are never used.' However, there are two vital questions to be asked before this conclusion can be proven:

Do I have access to all these nil-use titles?

Whatever version of the deal a library has, the usage statistics downloaded from the publisher will generally contain all titles available within all versions of its deal, with nil usage recorded against those not accessible. They will also include titles from any separately purchased backfile collection. The library has first to identify which titles are available within the deal purchased, and to separate out titles from the backfile collection.

Do all these nil-use titles actually exist?

Having eliminated all titles not available in the particular deal and all backfile collection titles, there may still be a large residue of titles with nil use recorded. By matching up titles in the usage statistics with those in the price list, it is possible to identify a number of titles that have no price recorded against them. These 'unpriced' titles may have notes against them in the price list marking them as 'ceased', 'transferred' 'title changed' or 'new acquisition'. They may simply not appear in the price list at all. By using the template to mark these 'unpriced' titles it is possible to get a more realistic view of which titles are genuinely unused and which are simply not available.

- Matching titles to subject categories: most publisher price lists contain an indication of a title's subject field. Sometimes this is a broad category such as 'HSS' or 'STM'; sometimes it is much narrower. For the project, both the broad and the narrow subject categories were given against each title. This enabled separate analysis of usage of HSS and STM titles, to see whether different patterns were apparent. It also allowed those libraries that

wished to do so to match up narrow subjects to their own faculty or school structure and do more detailed usage analysis by subject.

- Matching titles to subscriptions: in the current pricing models, the cost of maintaining existing subscriptions (whether print or online) is generally a significant factor in the total cost of the deal. Libraries were asked to mark up their subscribed titles on the template so that usage could be analysed by subscribed title and by 'unsubscribed' (those additional titles acquired as part of the deal). Looking separately at usage of subscribed and unsubscribed titles has the advantage of showing how well subscribed titles are being used, indicating a good choice of titles in a way that was never possible in the days of print journals. It also shows how much or how little the unsubscribed titles are being used, indicating the value of subscribing to the whole deal rather than having individual subscribed titles.

Under the deals negotiated for NESLi2 (National Electronic Site Licensing Initiative 2), the additional element for 'e-access' to all other titles in the deal is generally a much smaller sum than the cost of maintaining existing subscriptions. This is logical, as libraries could argue that these are the titles in which they have less interest. By relating the cost of subscriptions to the total requests for subscribed titles and the cost of e-access to the total requests for unsubscribed titles, it is possible to break down the average cost per request figure and show average costs per subscribed and unsubscribed titles separately. This latter figure will probably be much lower, depending on both the terms of the deal and the volume of use of unsubscribed titles and, if this is the case, it is a useful way of demonstrating the value of the deal.

A further measure developed for the NESLi2 project was the 'yield per £'. This makes use of the list prices added to the template to calculate the list price of all 'unsubscribed' titles in the deal. By dividing this sum by the e-access fee, the result or yield shows how many pounds' worth of titles can be bought for each pound of the e-access fee.

WHAT WERE THE PROBLEMS?

Having demonstrated the value of using the template, there were still some obstacles for libraries to overcome before they could start work on the actual analysis. The issues described here will be

familiar to anyone who has worked with usage statistics.

The titles don't match

As we had found with the NESLI2 study, a group of libraries will not necessarily have exactly the same number of titles in their usage statistics even when they all have the same deal. This was due mainly to the presence of 'unpriced' titles (see above) which appeared in a greater or lesser number in usage statistics. Providing a 'one size fits all' template was not therefore possible, and for libraries taking part the first hurdle was matching up their list of titles with those in the template. This proved either a cumbersome manual exercise or, for several, a first attempt at using Access or, if all else failed, a plea for help to the project leader.

What to do with the aggregators?

One of the mysteries of usage statistics is when to add statistics from gateways and aggregators such as Swetswise or Ingenta. The answer is 'It depends', but it is important to know when to add, as this could have a significant effect on total usage. Libraries using these gateways will also want to know what percentage of requests they are handling. For the project, libraries were advised when to add these additional statistics, using the table provided for the SCONUL statistical return.

What about backfiles?

As described above, titles from the backfile collection are generally included in the main list of titles for the usage statistics, although some publishers now also present them separately. If the library has not purchased the collection, the titles will show up as nil use; if it has, the usage is likely to be at a lower level than current titles and will distort the figures. Using the template, libraries could separate out backfiles and, where appropriate, analyse their usage separately in the report.

What are the subscribed titles?

Libraries themselves entered details of their subscribed titles on the template. This was often a more difficult job than it seemed. As we had found with the NESLI2 project, this information was not always readily available. The library would generally use a subscription agent and the costs of print subscriptions and online access to a deal would often span different departments.

Writing up the reports

Having presented above some of the issues involved in setting up the project, it is important

not to lose sight of its main aim, which related to both the collection and the analysis of usage statistics. This involved getting beyond the actual analysis and asking the question 'What does it all mean?'. However tantalising a set of statistics may be in their own right, if they are to be used by management they need to go through a process of distillation. Library managers are unlikely to be impressed by a host of spreadsheets and explanations of why one set of figures is incompatible with another, however important accuracy and consistency may be in the actual compilation of those statistics. They will be looking for some key pointers that they can use in budget discussions with senior management. Subject librarians will want to know how certain titles are being used within their area, where they may need to focus training and how they can demonstrate value for money to academic staff.

The report template was therefore designed to provide an opportunity for project members to analyse statistics and look for meanings and explanations in the results, and also to select key points that could be presented to management and lists of titles that could be used by subject librarians.

HOW DID IT GO?

Now that I have described in some detail all the difficulties involved in setting up the templates and writing the reports, it is a measure of the success of the project that, in spite of staff changes affecting several libraries during the year, all but one of the original fifteen libraries have stayed with the project, one backing out through pressure of work. To date, 40 publisher reports have been produced, covering five of the six publishers in the original list. Two libraries have done further work with Evidence Base, building on the usage data to look at evidence of impact.

One of the advantages of the project has been the opportunity offered for benchmarking results with other participants. Publisher and library confidentiality currently prevent individual results being shared more widely, but the agreement of both libraries and publishers for the sharing of information within the project itself has been an important factor. These publisher reports are all available to project members in a secure file area, where they can be viewed or downloaded for benchmarking. One publisher report with summary tables showing key indicators for all participating libraries has been completed and others are being prepared. Libraries are still actively work-

ing to finish their sets of reports and are being encouraged also to produce their own summary tables showing key results across all the deals.

Evaluations received to date have been encouraging. Participants report:

'We have found the project interesting and have enjoyed participating. It has certainly provided us with some new techniques to employ when conducting our own value for money analysis.'

'A much better understanding of the complexities of compiling accurate usage data.'

'Better understanding of the ratio between cost and usage of our subscribed and unsubscribed titles.'

'An awareness of how to analyse the deals by looking at more than usage figures.'

'A very worthwhile project.'

'A good methodological starting point.'

'A framework and timescale to work to and get us started on doing something we knew was needed.'

'I have found the exercise useful and it has given me insights into how our collections are used.'

The challenges of manipulating the data using advanced features of Excel and, in particular, learning Access were mentioned by several participants. Finding time to set aside for the work was inevitably a constraint.

Some libraries have already presented completed reports to various groups and received good feedback:

'The completed reports have been very useful for the detailed light they shed on particular agreements.'

'Displayed to management to show how much value for money or otherwise we were getting or deals – very positive feedback of possible usefulness.'

'Electronic resources steering group have found these useful in considering "value for money" in bundled deals.'

All plan to continue with the process of analysing deals, either on their own using the methodology

developed for the project or by becoming part of a future group in this project.

FUTURE PLANS

Following the successful completion of this first round, Evidence Base are looking to offer this project again for a further period, assisting libraries to analyse their usage for 2006 and, where appropriate, to track trends over the previous two years covered by this project (2004–5).

It is hoped that some of the existing libraries will stay on for a further year and be joined by a new group. As well as sharing the expertise that libraries have built up, this approach will enable us to build up the benchmarking with more libraries. We hope to expand the number of publishers and to move to other formats such as e-books.

While libraries have gained in knowledge of how to analyse and present usage statistics, it is recognised that this is a time-intensive process and that some libraries may prefer to outsource the work. For this reason, Evidence Base will also be offering to produce the reports for any libraries that prefer this approach. Costs and details are currently available on request.

REFERENCES

- 1 Conyers, A (2004). 'E-measures: ready for the count?', *SCONUL Focus*, 31, pp 53–6. Available at <http://www.sconul.ac.uk/publications/newsletter/31/18.pdf>
- 2 Conyers, A. and Dalton, P. (2005), 'NESLi2 analysis of usage statistics: summary report'. Available at <http://www.ebase.uce.ac.uk/docs/jiscnesli2summaryeb.pdf>
- 3 COUNTER. (2005) *COUNTER code of practice. Journals and databases: release 2*. Available at http://www.projectcounter.org/r2/COUNTER_COP_Release_2.pdf

Anyone interested in finding out more about the *Analysing Publisher Deals* project can contact Angela Conyers at Evidence Base (angela.conyers@uce.ac.uk).