

ELECTRONIC JOURNAL USAGE DATA WITHIN NESLI

Ross MacIntyre

Paper delivered at the UKSG Seminar – Usage Data for E-Collections, 7th June 2000

The presentation concerns the production of usage data within NESLI. This involves MIMAS working with an aggregator, SwetsBlackwell, to produce usage reports requested by the initiative's Steering Committee.

NESLI (National Electronic Site Licence Initiative) was established in 1998 by the UK Higher Education Funding Councils' Joint Information Systems Committee (JISC). The overall aim of NESLI is to facilitate and promote the use of e-journals in UK Higher Education Institutions (HEIs). NESLI might be described as being a (voluntary) virtual consortia of UK HEIs and specifically the only one approved by JISC to negotiate for full-text e-journal content on its behalf¹.

Following a formal tendering process, a consortium of Swets & Zeitlinger and University of Manchester was appointed to act as the Managing Agent (MA) for the initiative. The MA reports to a Steering Committee composed mainly of senior UK academic librarians. The MA performs the day-to-day service operations: publisher negotiation, communicating offers, taking orders, service delivery (where orders have been placed with SwetsBlackwell), support and a small amount of service development.

Status of NESLI

NESLI is in its second year of three. At present (June 2000) there are twelve offers on the table: Blackwell Publishers; Blackwell Science/Munksgaard; Elsevier (Science Direct); MCB University Press; Kluwer; Project MUSE (from Johns Hopkins University Press); Mary Ann Liebert; National Research Council Canada; Oxford University Press; American Chemical Society; Internet Archaeology; and Academic Press. Obviously many publishers have been, and continue to be, in discussion with the MA and offers are at various stages. Together the above represent over two and a half thousand journal titles, available in electronic form. Over eighty-five sites are subscribing to one or more NESLI deals.

Ross MacIntyre, Senior Project Manager, MIMAS, Manchester Computing, University of Manchester.

Usage Statistics

NESLI usage statistics are collected and disseminated by MIMAS (Manchester InforMation and Associated Services), a section within Manchester Computing at the University of Manchester, home to one of the three National Data Centres. MIMAS hosts data and applications principally for the UK Higher Education community, though users also include the research councils, further education and other European academic institutions.

When an HEI selects SwetsNet² as the delivery service for access to NESLI e-journals, librarians within the institution have access to usage statistics on-line via an administrative interface. Statistics are available for access at three levels: Table of Contents (ToC); Abstract and Full-text. These can be viewed grouped by journal title, issue and by month, in hierarchical fashion. Figure 1 shows a ‘dummy’ administrative account used by SwetsBlackwell for demonstration purposes. Displayed is a list of full-text subscriptions across all journal subject areas, showing all ToC, Abstract and Full-text accesses since the subscription began.

If a user clicks on a specific title, the information in Figure 2 is displayed showing accesses per issue.

Figure 1



Figure 2

Clicking on the ‘statistics per month’ link will show when these accesses took place – see Figure 3.

Data Gathering

SwetsNet is a web-enabled interface to an application running on a server at Swets & Zeitlinger’s headquarters in Lisse, The Netherlands. The application produces various log files, created from individual transactions, including one relating to usage. During 1999, the contents of the log file was expanded and now consist of:

- A - Country Code
- B - Customer Name
- C - Agency Code
- D - Sub Department Code

- E - Account ID (which is what MIMAS converts the ATHENS institutional identifier to so that SwetsNet does not have to change to support ATHENS)
- F - Year
- G - Month
- H - 080-Code. The ‘content information field’, containing the following items:
 - I - Publisher ID
 - J - Publisher Name
 - K - Title ID
 - L - Title Name
 - M - ToC Hits
 - N - Abstract Hits

- O - Full-text Hits
- P - Subject
- Q - Account Code (C=customer, N=NESLI). This data is used to tailor the display for NESLI customers
- R - ISSN Electronic Version
- S - ISSN Print Version (Note that although SwetsNet differentiates between the electronic and print version, either ISSN can be quoted when linking, since linking is only relevant to the electronic version)

The five tables consist of:

- the raw data itself;
- data sorted and totalled by customer (institution);
- data sorted and totalled by publishers;
- data sorted and totalled by journal title;
- data sorted and totalled by subject.

MIMAS take the customer, journal title and publisher data and perform some minor processing to produce the final versions of the use data spreadsheets requested by the NESLI Steering Committee.



Figure 3

Processing the Data

The SwetsNet log file is large as it contains data relating to the company’s global user base, so a NESLI-specific extract is required. Once a month, the log file is ‘swapped’ on the main application. SwetsBlackwell in the UK download the ‘swapped’ log file and load it into an Access database.

A new table is created containing only data relating to NESLI customers. From this smaller table, a further five tables are created and exported as Excel spreadsheets and passed to MIMAS. From start to finish the creation of the spreadsheets represents around two hours work each month for SwetsBlackwell.

Steering Committee Reports

Included below are some extracts from the Steering Committee reports. Certain information has (deliberately) not been included in the extract to avoid any confidentiality issues.

Figure 4 below shows one month’s accesses per institution. From the report it can be seen that, for example, in Leicester there were almost twice as many Full-text accesses as ToC, whereas in Durham, the opposite is true. In this paper I will not speculate on what this might indicate.

Figure 4

	March		
	TOCs	Abstracts	Fulltexts
Aberdeen, University of	960	67	561
Bangor, University College of Wales	617	260	502
Birmingham, University of	1420	214	1160
Bournemouth University	438	121	1520
Bradford, University of	462	77	332
Cambridge, University of	682	71	777
Cardiff University	304	43	199
Cranfield University	368	58	794
Dundee, University of	324	58	237
Durham University	1530	117	699
Edinburgh, University of	452	28	356
Glamorgan, University of	269	18	401
Glasgow Caledonian University	198	20	45
Glasgow, University of	1156	101	681
Hull, University of	268	90	344
Imperial College of Science, Technology and Medicine	1026	136	1151
Keele University	162	23	72
Kingston University	69	32	44
Leicester, University of	773	206	1336
Liverpool, University of	307	51	233
Newcastle upon Tyne, University of	820	64	454
Northampton University College	344	123	322
Open University	1066	288	1197
Paisley, University of	318	38	153
Queen’s University Belfast	2110	358	1504

Figure 5: 2000 NESLI Usage Stats – by publisher

Publisher	March			
	No of titles	TOCs	Abstracts	Fulltexts
Blackwell Publishers	219	3368	620	2033
Blackwell Science Ltd	220	14563	2355	14240
Project MUSE	112	861	9	290
Kluwer Academic Publishers Group	609	3336	510	2565
Mary Ann Liebert	15	38	0	5
MCB University Press Ltd	114	1048	298	1258
National Research Council, Canada	12	17	0	0
	1301	23231	3792	20391

Publisher reports

Figure 5 shows the total usage for March 2000 for a subset of NESLI publishers.

It would appear that, comparatively speaking, Blackwell Science titles were being downloaded most frequently. However, it might also be the case that the downloads were all of just one acclaimed article.

The journal title report in Figure 6 shows the number of Full Text, ToC and Abstract accesses to specific journals. It is interesting to speculate why so many accesses stop at ToC level for the *Journal of Business Finance and Accounting*, whereas this was not the case for *Journal of Business and Industrial Marketing*? Does this indicate that these articles have been found by searching rather than 'traditional' hierarchical browsing? This is one small example of the difficulties of interpreting usage statistics.

Completeness of the Data

The statistics accurately reflect all accesses to NESLI material obtained via SwetsNet, whether using ATHENS, IP or SwetsNet login. They also include the IP accesses representing direct linking, for example, from library OPAC or e-journal page. The links tend

to be at journal title level, though SwetsNet supports accesses at issue and article level.

However, NESLI-arranged offers do not compel institutions to use SwetsNet as the delivery service. Many of the deals allow access either directly to the publishers or via their agent, for example, ingenta. So the coverage is partial in that respect. It cannot include accesses via these other routes. At present some publishers do not make statistics available at all. In the case of ingenta, they cannot pass usage statistics to the library, as they work for the publishers, who claim that they 'own' the usage data.

Secondly, even when material is accessed via SwetsNet, it is possible to 'lose' the user. Consider the case where a publisher insists that when an article is requested, the user should not be passed the file immediately. Rather, they should be shown the publisher's own service, featuring, what the publishers feels are, 'value-added' functions. To the user this really means

Figure 6: 2000 NESLI Usage Stats – by journal title

Journal title	Publisher	March		
		TOCs	Abstracts	Fulltexts
Journal for General Philosophy of Science	Kluwer Academic Publishers Group	9	4	6
Journal for the Scientific Study of Religion	Blackwell Publishers	5	0	0
Journal for the Theory of Social Behaviour	Blackwell Publishers	13	0	0
Journal of Advanced Nursing	Blackwell Science Ltd	1026	323	1973
Journal of Analytical Chemistry	Plenum Publishing Corporation	1	0	0
Journal of Animal Breeding and Genetics	Blackwell Science Ltd	3	0	1
Journal of Animal Ecology	Blackwell Science Ltd	78	21	58
Journal of Antimicrobial Chemotherapy	Oxford University Press	1	0	0
Journal of Applied Ecology	Blackwell Science Ltd	162	11	98
Journal of Applied Electrochemistry	Kluwer Academic Publishers Group	19	1	20
Journal of Applied Microbiology	Blackwell Science Ltd	202	34	265
Journal of Asian American Studies	Johns Hopkins University Press	8	0	1
Journal of Banking and Finance	Elsevier Science Ltd	1	0	0
Journal of Bioeconomics	Kluwer Academic Publishers Group	3	0	0
Journal of Bioenergetics and Biomembranes	Plenum Publishing Corporation	6	0	1
Journal of Business and Industrial Marketing	MCB University Press Ltd	16	25	89
Journal of Business Ethics	Kluwer Academic Publishers Group	25	6	13
Journal of Business Finance and Accounting	Blackwell Publishers	132	32	32
Journal of Interdisciplinary History	MIT Press	0	1	6
Journal of Interferon and Cytokine Research	Mary Ann Liebert	1	0	2
Journal of Peptide Research	Munksgaard	6	0	3

they are shown the abstract again. (This is commonly called "branding" by publishers, but is usually described with different terminology by the end-users, who have already indicated they wanted the full-text article.) Once within this different service delivery environment, further access data is lost to the SwetsNet log file.

Consumers of the Usage Data

Currently, HEIs have access to NESLI statistics via the SwetsNet on-line interface. There has been some feedback requesting additional data and alternative presentations or views of the data, for example, sorting in order of Full-text accesses or in reverse order. These requests have to be dealt with by SwetsBlackwell. Only one site has requested raw data files, which MIMAS has provided. This is more likely to be down to resource and time constraints than lack of interest.

Usage data is also extremely valuable to the MA when (re-) negotiating deals with publishers. Statistics illustrating use, or lack of, offers leverage. The NESLI support officer at MIMAS does occasionally receive enquiries about usage, but the data is primarily used to compile internal and external reports.

NESLI Steering Committee

It should be stressed that the NESLI Steering Committee works very closely with the MA and every potential offer is scrutinised by them. The committee is very interested in testing new economic models and hypotheses and utilising usage data to inform the decision making process. What follows is a series of comments from members of the Steering Committee relating to usage data.

"People have now got a chance to get statistics on usage they have not had before. This applies to both producers and consumers and usage data should be available to both to guide sensible purchasing."

"There is some frustration with the 'all or nothing' pricing model which forms the basis of so many of the offers tabled, all seemingly requiring the libraries to spend more, so as to safeguard

existing revenue streams. Maybe pricing models should accommodate holdings versus access decisions and include both access to whole titles and document delivery, or document delivery exclusively. Why base pricing on print anyway?"

"Data is often collected for the purpose of testing a hypothesis. Here is a hypothesis which we need to test, if we can, but have never been able to test before. A large proportion of journal articles are never read by anyone, though people may look at the title and possibly the abstract as well. The proportion of articles falling into this category may even exceed fifty per cent. They are published not because they need to be disseminated to a wide audience but in order to meet other criteria, such as personal career development or (in the UK) the RAE."

"A journal is a bundle of articles in which the articles people do want to read, sell the articles which people don't want to read (as well as selling themselves). Some journals contain no articles that people want to read but are bought just in case someone might want to read one of the articles."

Conclusion

NESLI usage data ought to begin to give us some indication whether some or any of the above hypotheses and statements are true. If true, a great deal of money is being wasted printing onto paper, in many copies, articles which people do not wish to read - although there are other grounds for making it clear in some way that these articles do exist and can be consulted if necessary. If true, the present academic journals industry is a gross waste of resources and the sooner a different model is introduced, the better.

From my own technician's perspective, it is clear that there are shortcomings in usage statistic collection, but refusing to supply usage data on the basis of people's possible reaction to seeing zeros in a particular column is naive and unacceptable.

References

1. The NESLI web site: <http://www.nesli.ac.uk>
2. SwetsNet was rebranded in July 2000 as SwetsnetNavigator