

**David Nicholas, Paul Huntington, Ian Rowlands,
Tom Dobrowolski and H Jamali**

CIBER, University College London, UK

1 Abstract

The paper represents the release of log data from the SuperBook Project, which sought to investigate the impact of the introduction of e-books to the University College London (UCL) scholarly community using an evidence-based methodology. It presents a deep log analysis of e-books usage of Oxford Scholarship Online (OSO), one of three e-book packages introduced to UCL. A range of analyses was conducted in order to evaluate use and patterns of information-seeking behaviour. Use was measured over a three-month period and in the following ways:

- a) the number of pages viewed;
- b) the number of sessions conducted;
- c) the number of views in a session;
- d) the time spent online; and
- e) the number of pages printed.

In all, 1277 sessions were conducted and 10,678 pages viewed during January to March 2007. Nearly three-quarters of sessions saw content viewed and 17 per cent recorded a page printed via the OSO print routine. The findings suggest that e-books will become very popular and that e-book information-seeking behaviour differs from that associated with e-journals.

2 Introduction

Research into how digital resources are used within the academy has largely focused on e-journals. However, the virtual scholar uses a much wider range of digitally delivered content to achieve their research, teaching and learning goals, and there is a clear danger that scholarly information seeking is defined or coloured by what has been witnessed in the e-journal environment. As a first step towards obtaining a more rounded picture of how digital resources are used CIBER's Virtual Scholar Programme is currently subjecting e-books to the same robust methods that has been previously reserved for journals. In recognition of the belief that e-books have the potential to transform the scholarly environment, perhaps to a greater extent than e-journals, the project has been named SuperBook. The reasons e-books could have a greater scholarly impact are:

- 1) They are more likely to be of value to the biggest scholarly community, students, who typically report on the difficulties of obtaining key textbooks, and to whom journals are of limited value.
- 2) Even for faculty members we might expect to find a big impact because books have never been as easily accessible as journals and for certain fields, like the arts and humanities they are the prime source of research.
- 3) They will be more suitable than e-journals for the further education sector.

While there has been much talk about the great market potential for e-books there has been little in the way of robust, evidence-based user studies which would support or disprove the contention. There is of course a relative abundance of self-report studies (e.g. Armstrong, Edwards

and Lonsdale, 2002; Chu, 2003; Langston, 2003; Levine-Clark, 2006) but there are clearly considerable dangers in asking users to comment on the impact and future of a newly emergent technology of which they have little experience.

3 Aims and scope

SuperBook was an action research study¹, which involved 'dropping' more than 3000 selected e-books from OUP (Oxford Scholarship Online), Wiley (Interscience) and Taylor and Francis into the UCL information environment and then assessing, by means of deep log analysis, what happened as a result. What was being created was an e-book observatory in which behaviour could be observed and changes introduced and then evaluated. This paper concentrates on one of the e-book collections, Oxford Scholarship Online.

Conventional transactional log analysis of the kind provided by publishers in COUNTER-compliant form for libraries can only provide periodic, broad and shallow indicators of activity, whereas deep log analysis (DLA) provides a detailed and real-time assessment of the information-seeking behaviour of user communities and these data can be used to help determine impacts through further qualitative means. DLA involves the processing of huge volumes of usage and search data as provided in the raw transactional logs of publishers and then relating this to user demographics to provide a whole range of evidence-based user portraits – hence the word 'deep'. This information then provides the foundation for follow-up user surveys and interviews (to be conducted in the autumn of 2007).

¹ Funded by Wiley and Emerald

The aim of this paper is to demonstrate what can be squeezed from the e-logs in order to furnish a whole range of metrics that could be used to profile the information-seeking behaviour of virtual scholars using e-books. Wherever possible, information-seeking behaviour has been related to user characteristics.

Usage was measured in number of ways to obtain a more accurate and comprehensive picture:

1. Number of pages reviewed.
2. Number of sessions conducted.
3. Number of views per session.
4. Time spent viewing a page.
5. Duration of a session.
6. Number of pages printed.

Use was related to various user characteristics:

1. Subject I: defined by subject of book viewed.
2. Subject II: defined by sub-network label of server used.
3. Geographical location of user: on-site/off-site.
4. Academic status: staff, student, defined by sub-network used.

The following specific information-seeking characteristics were evaluated:

1. Referrer link used (e.g. Google).
2. Form of navigation to content adopted.
3. Age of book viewed.
4. Individual e-book titles used.
5. Number of e-books used.
6. Scatter of usage over titles available.
7. Whether catalogued or not.

OSO was introduced to UCL in November 2006. Two studies of OSO were conducted. The first, a pilot study, covered the period November to December 2006. The purpose of this was to obtain a greater understanding of the e-book logs. This was necessary because they are different to those of e-journals and it was necessary to fine-tune analyses after consultations with both OUP and UCL librarians. The second study, reported here, informed by the first, covered the period January to March 2007. This was also a period when the profile of e-books was raised, largely by giving it greater prominence on the library web page and we wished to investigate the impact.

4 Background

Oxford Scholarship Online (OSO) is a cross-searchable library containing the full text of over 1,200 books on economics, philosophy, political science, and religion. Specially commissioned (author) abstracts and keywords are provided at book and, unusually, chapter level. In the case of the OSO collection the e-books are e-monographs rather than e-textbooks, so it would have been expected to appeal to all humanities and social science scholars – staff and students alike.

Figure 1:
Screenshot of Oxford Scholarship Online homepage



5 Methods

The following box gives an example of a line from the OUP e-book transactional server log file. The internet protocol address has been made anonymous by substituting xxx's for last two sets of digits.

```
144.82.xxx.xx 59D423E0E0F85EDEA6FCEF3D1427C706 5204
[03/Nov/2006:14:39:47 +0000] "GET /oso/private/content/
economicsfinance/0199271488/p015.html HTTP/1.1" 200
55540 "http://www.oxfordscholarship.com/oso/login.jsp?
errorMessage=Youper cent20doper cent20notper cent20
haveper cent20accessper cent20to per cent20anyper cent
20titles&forward=http://www.oxfordscholarship.com/oso/
private/content/economicsfinance/0199271488/p015.html"
"/oso/private/content/economicsfinance/0199271488/
p015.html" "Mozilla/5.0 (Windows; U; Windows NT 5.0; en-GB;
rv:1.8.0.7) Gecko/20060909 Firefox/1.5.0.7"
```

The first field 144.82.xxx.xx records the internet protocol (IP) address. The IP number is a numeric address that is allocated to users connecting to the internet. The field 59D423E0E0F85EDEA6FCEF3D1427C706 is a unique session ID and is set in a cookie once authentication has occurred. The UCL account number is 5204. The date and time is given in the date and time stamp, and records when the file was uploaded to the client's computer.

The field beginning with GET is the URL of the actual content along with the directory name of the file uploaded. The directory is structured around the ISBN for the book (e.g. 0198293542) which is present for every specific book-page on the site. A database of OUP ISBN codes was supplied and book detail fields added to the dataset. The p015.html is the number of the specific HTML page sent. This is not the print page number as the HTML page contains around five printable pages. The site has a distinction between /public and /private. The latter is everything behind access control, namely the full text of the books. The remainder of the site, particularly the table of contents (TOC) pages, are on /public pages and freely accessed/indexed by Google.

The logs further record the status of the upload, the number of bytes sent, the referrer link and the browser details of the client machine.

Logs were subject to standard deep log techniques and processed by SPSS.

6 Results

Use can be measured in a number of ways:

- the number of pages viewed;
- the number of sessions conducted;
- the number of views in a session, a hybrid metric;
- the time spent online;
- the number of pages printed, which is also an important satisfaction metric.

These metrics are used to investigate diversity and to create profiles of various user communities.

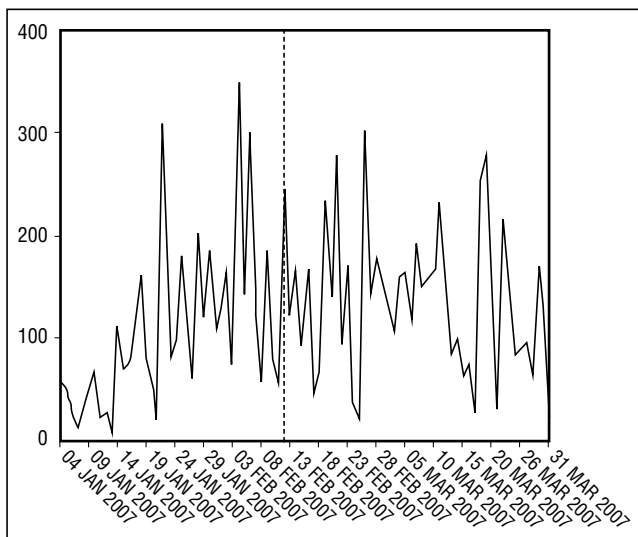
6.1 Usage

6.1.1 Pages viewed

What actually constitutes a 'page' is worth spelling out. In the case of OUP e-books a page might be the homepage, a search page, list of books, abstract or actual book content. In the case of content each page is an HTML page which is the equivalent of five print pages.

The average number of page views per day was 124, which probably shows the relative novelty of e-books in the UCL environment (Figure 2). During the period the service was moved to a more prominent part of the UCL library services website to see what impact this would have, and this occurred during the second week of February (marked by a dotted line). The change, surprisingly, did not appear to have an impact on daily use, which is characterised by volatility and the fact that there is no rising trend. This could be due to the fact that we are dealing partly with students here whose e-book attention span lasts only as long as the module (ten weeks).

Figure 2:
Frequency of page views by day,
January to March 2007



In terms of day of week, Mondays attracted the greatest use, accounting for 20 per cent of weekly page views. Saturdays recorded the lowest figure of just 7 per cent. Unexpectedly, Sundays recorded a relatively high level of use, 14 per cent of page views. It should be noted that those who attempted to access the actual content pages outside UCL needed to log in to the OSO site using their Athens password if they came in externally via a search engine or enter their UCL id/passwords if they entered the service from within the UCL website.

6.1.2 Sessions conducted

On average 15 sessions per day were made. The average number of sessions per day was about 8 or 9 in the first three weeks of January rising to 17 sessions per day by mid-March and daily use remained at about this level. There was a peak in session usage around 19 March when the number of sessions reached 60.

6.1.3 Number of page views in a session (site penetration)

One in five sessions conducted (21 per cent) saw just one page viewed, a further 24 per cent viewed two to three pages, 32 per cent viewed four to ten, 13 per cent viewed between 11 and 20, and 10 per cent viewed over 20 pages. Comparing these findings with those found for journal usage produces interesting results. Thus, in the case of OhioLINK, the figures were respectively, 18 per cent, 31 per cent, 35 per cent, 10 per cent and 6 per cent. The basic pattern, with the proportion increasing and peaking at between four and 10 page views per session and then declining is the same. Where the chief differences lie is:

- in the proportion of sessions viewing a large number of pages (11 or more) – for journals the proportion was 16 per cent but 23 per cent the case for books; and
- that there is a much higher proportion of journal sessions viewing two to three pages – 31 per cent compared to 24 per cent for books.

Together this would suggest that people were viewing more e-book pages in a session. This could be due to a number of reasons:

- Viewing e-books is a relatively novel experience and this results in people looking around more.
- The page actually refers to an HTML full-text page which, in the context of a content page, contains five print numbered pages and users will have to view a number of pages to view a chapter and this scrolling through pages to view a chapter increases the number of views in a session.
- The e-book user population consists of a higher proportion of students and it reflects their different information-seeking behaviour.

6.1.4 View time

The average page view time was about 14 seconds. It should be pointed out that the average value here is not a good indicator of online activity as a small percentage of use will represent people spending a lot of time reading content online; that is, e-books are largely designed to be read online. Like page view time, session time is skewed. On average, sessions lasted over three and half minutes. The median session length is estimated was three minutes, which puts into perspective the relatively short time people spend online, on any single site. It was found that 14 per cent of sessions had a session length of over 15 minutes and 51 per cent of sessions lasted three minutes or longer.

6.1.5 Printing of pages

The principal sign of satisfaction or outcome from using OSO is the viewing and printing of e-book pages. In this respect fewer than one in three sessions did not lead to an outcome; that is, these session just viewed pages other than e-book content, 70 per cent of sessions saw content pages viewed and 17 per cent recorded an e-book page printed via the OUP print routine. The print option here is a specific OUP page that the user requests, the client might also print using the browser option but this was not recorded in the logs.

6.2 Use by various e-book and user characteristics

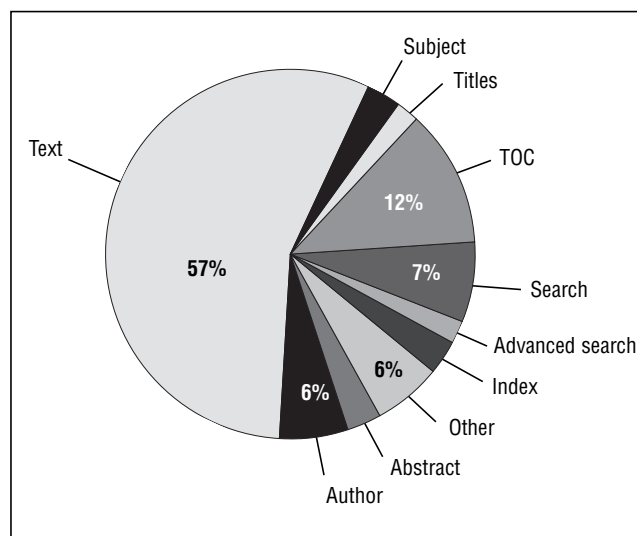
Deep log analysis enables the log data to be mined in order to establish differences in patterns of use between different types of user and e-books. In this connection the following section provides an analysis of:

1. Use by type of page viewed. When users arrive at the site they could view a range of pages and this provides an insight into the kinds of activities undertaken, and in some cases, provides an indication of how satisfied users were with what they saw.
2. Use by number and titles of e-books viewed.
3. Use by the physical location of the user (whether they accessed OSO on-site or off-site). The on-campus versus off-campus analysis is clearly an interesting one, especially for librarians who might be worried by users, especially students, not using the physical library.
4. Use by referrer link used (website from which the user arrived at the OSO site).
5. Use by method of navigation employed to find content. The navigation variable links two access variables: access as defined by referrer link and searching/browsing information as summarised from download items viewed.
6. Use by sub-network label of computer used (provides academic status and subject field of user). Sub-network labels are non-mandatory labels assigned by system controllers to parts of the computer network. Labels may or may not represent actual user locations but in general system controllers will label networks meaningfully (e.g. 'phil' corresponds to computers located in the philosophy faculty). Faculty networks will be used by faculty and research staff. Some networks will be used predominately by students and in this study it was found that the sub-network label 'hor' is located in the halls of residence and hence mainly used by students.
7. Use by subject of e-book.
8. Use by age of e-book.

6.2.1 Type of page viewed

Fifty-seven per cent of views were to actual e-book content; that is, full-text pages (Figure 3). It is worth repeating that these are not a single page view but a chunk of printable pages, typically of four to five print pages in length. The fact that 43 per cent were not content pages demonstrates that finding your way around an e-book site really is a significant activity, and OSO provides a wide range of means by which this can be done. Views to other pages included: table of contents, which also includes an abstract (12 per cent), author index (6 per cent) and subject index (3 per cent). (The latter being pages which are essentially menus or lists that allow users to browse through the database.) Abstracts only accounted for 3 per cent of page views but abstracts were also on view in the table of contents page and these abstract pages were actually chapter abstracts, not whole book ones. There were also two search options: quick search (7 per cent) and advanced search (1 per cent). With quick search you just type a word and go while advanced search gives users more options to construct complex queries and to refine their searches.

Figure 3:
Percentage frequency of type of page viewed



It was no surprise that content (full-text) pages recorded the longest average (median) view time of 15 seconds. Of course, this is not enough time to read the pages, just enough time to assess their relevance. Of the content pages, about 40 per cent were viewed for 10 seconds or less, 31 per cent for between 10 seconds and half a minute, 12 per cent for between 31 and 59 seconds, 10 per cent viewed a page from between one and two minutes, and 7 per cent viewed a page for over two minutes. The table of contents recorded a view time of 11 seconds, which suggests that people might require help in orientating themselves in a new and novel information environment. Many of the search approaches available take users directly to the ToC page, so a degree of initial orientation is to be expected.

Off-site users adopted a more direct approach and viewed a far greater proportion of content pages: 72 per cent of views were to content pages. By contrast, the equivalent figure for on-site users was 54 per cent. It is hypothesised that off-site users perceived their access to be less permanent or less reliable and thus spent less time searching around, being keen to obtain all the required page information while they could, just in case they had access problems later. UCL site users on the other hand felt more secure, viewing the access as institutional and therefore their use could be safely spread over a number of visits. This information-seeking hypothesis could be termed 'access expectation'.

6.2.2 Number and titles of e-books viewed

Oxford Scholarship Online consists of over 1,200 e-books. Just two of the 1222 accounted for over 12 per cent of the page views: *Justice Posterity and the Environment* and *Justice Beyond Borders*.

Under a third (30 per cent) of sessions did not view any book title at all. Users not viewing an e-book title were viewing a menu page only and one hypothesis is that users were verifying availability and checking that they could access the service. Nearly half (49 per cent) of all sessions just viewed one book, 16 per cent viewed two or three e-books and 6 per cent viewed more than four titles in a session.

There were big differences according to:

- a) the physical location of user;
- b) referrer link used;
- c) method of navigation employed; and
- d) sub-network label of computer used.

Those people accessing from a non-UCL domain address were more likely to view an e-book in a session. Just 9 per cent of this group did not view a title compared to 40 per cent of UCL domain sessions which did not view a title. This is a huge difference and, again, this may reflect access expectation on the part of the users. Users within UCL may expect the service to be maintained hence they may be just taking a look to see that it is there with the future intention to access the material when they want to use it.

Search engine users viewed the most titles in a session, probably downloading more as a result of the novelty of finding the material free or having a lower access expectation. In other words, they assumed it would be better to download straightaway, rather than postpone the download. Those coming in via Oxford Scholarship Online were most likely to view just one title with about two thirds (64 per cent) doing so. Oddly, half (52 per cent) of UCL users did not view a title, again perhaps these users postponed their access expecting to return to what they must see as an institutional service?

Confirming previous research (Nicholas, Huntington, Jamali and Tenopir 2006a) those people navigating using the menu as well as the search option viewed more titles in a session: 41 per cent viewed two or more titles in a session compared to 28 per cent of those session users just using menus.

Nearly half (47 per cent) of sub-network sessions labelled gene (biomedical, genetics and history of science users) and 53 per cent of philosophy ones viewed two or more titles; furthermore, about 7 per cent of users on each of these networks viewed ten or more titles. Both of these are faculty networks and the usage recorded could be staff exploring what is there. Roam, a wireless network favoured by students, and UCL, used by both staff and students, were least likely to view more than one title with, respectively, 26 per cent and 15 per cent viewing just two or more titles.

6.2.3 Physical location of user

Over two-thirds of e-book usage took place within UCL but, of course, not necessarily in the library. However, the proportion varies quite considerably during the week. Thus about a third of off-site usage occurred on a Sunday. The percentage distribution of UCL-originated use across days showed that weekend usage was just 5 per cent, while each weekday accounted for about 15 per cent of usage. Mondays accounted for a surprisingly high 25 per cent, suggesting that a class related to e-book usage might have been held on the day.

In terms of views in a session – a possible ‘busyness’ or interest indicator – off-site users were more likely to view fewer pages with 52 per cent of sessions seeing three or fewer pages viewed. The equivalent figure for on-site users was 42 per cent: a real difference that might suggest more serious searching by those on campus, perhaps in the library.

In terms of session time, off-site users recorded a longer session length; well over a third (37 per cent) recorded a session length of over seven minutes as compared to 29 per cent for UCL domain users.

6.2.4 Referrer link used

The referrer link in the logs denotes the last site visited before accessing the OSO service. Fewer than a third (30 per cent) came in via a UCL site link, 42 per cent entered via the Oxford Scholarship site, 17 per cent entered via a search engine, and 10 per cent came in by other links. The large number of sessions coming from Oxford Scholarship site was unexpected. It is of interest to note, and this could be the beginning of a trend, two sessions connected to the service via Facebook (uclac.facebook.com).

Those accessing the service via Oxford Scholarship Online recorded the longest session time and 41 per cent recorded session times in excess of seven minutes. Interestingly this group were found most likely to view just one title perhaps these users were reading or working with the material while online. Given the fact that half the UCL sessions did not view a title it is perhaps surprising that this group still recorded quite lengthy sessions, with 47 per cent of sessions lasting three minutes or longer. Again, perhaps, these users were reading or working online. It should be pointed out that the way OUP presents e-books (in four to five page HTML chunks) means that it is convenient for users to use and work with the material while online, and it does take time to cycle through the content.

In terms of the number of pages viewed in a session, again, surprisingly, those accessing via a search engine were most likely to make more views in a session. Thus 68 per cent of search engine users viewed more than four pages in a session, whereas the equivalent figure for UCL domain users was 52 per cent. This is unusual, in that search engine users were known from previous studies to view the least number of pages. Perhaps, in the relatively early days of e-book access, search engine users did not expect to find the material free, they had a negative access expectation, and were eager to grab the opportunity to view or squirrel away more pages. Users coming in via Oxford Scholarship Online and UCL recorded fewer views in a session. Perhaps their relative access expectation was positive, that is they expected the service to be around for the foreseeable future and there was no pressure to download.

6.2.5 Method of navigation adopted

Navigation refers to the combination of referrer link used and searching/browsing method employed in order to access content. In the case of referrer link the user might jump to content from a link either from the UCL site or a search engine. However, users can also access content browsing via site menu or search using the in-site search engine. Seven per cent of sessions saw the search facility used, 34 per cent used only menus, 20 per cent used a combination of menu and on-site searching, 4 per cent used only a UCL link without apparently using any on-site menus or searching, 1 per cent used only a search engine link and 35 per cent were not identified. The size of this last grouping does limit the scope of this analysis and indicates the need for further research.

Those sessions when both menus and search screens were viewed recorded longer session times. These sessions viewed more titles and recorded a greater number of views. As expected, those using both the internal search option and menu viewed more pages in a session compared to other groups; 49 per cent of this group viewed eleven or more pages in a session.

Focus on search engine users

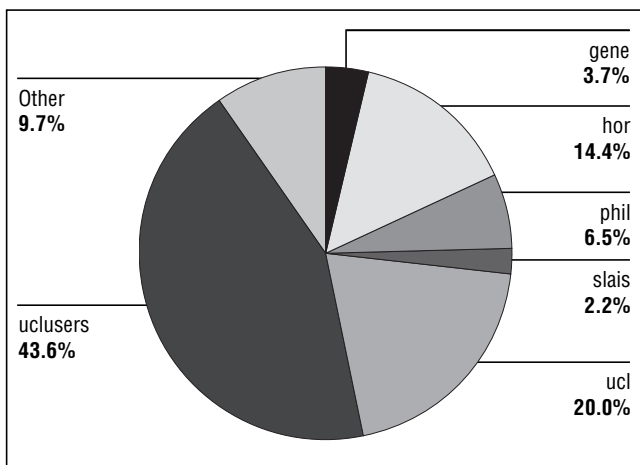
Scholars could use a world wide web (external) search facility to locate OSO e-books or the (internal) search facility. In general those people employing the internal facility used an average of 2.1 words when composing a search expression as opposed to 3.3 words in the case of external search engine users. There were also differences between the subjects of the book viewed. Thus, those searching externally and finding political science titles and economics and finance titles, used one more word in their search expressions when compared to those searching for philosophy or religion titles: four words compared to three words. With regard to the sub-network, Roam users used the greatest number of words (five) when formulating an external (WWW) search expression.

6.2.6 Sub-network label of computer used

Logs record use by computers and only provide a trace of the user's identity through the IP address provided. In this analysis we attempt to maximise this user trace. A sub-network analysis attempts to identify from the IP address where in UCL people were searching from and provides data about the subject background and the academic status of UCL users (staff, students). However, it should be stated that is not easy to say anything for certain about either location or users from any IP. Generally, IPs are allocated in blocks of 4, 8, 16, 32, 64 and so on, addresses, and these usually correspond to a physical switch or router and thus imply a physical location. But switches can be programmed, so there is no certainty that the logical tree structure of the IP address space will resolve into a corresponding physical tree. DNS names link to individual IPs so names even on a small subnet may not all be for the same department or location. Even when a department or location is implied by the DNS name it does not say anything about who is using the machine or if the use of that connection is limited to users within a department.

Three quarters of usage originated from UCL, for these users sub-network names were available; the most important by usage is given in Figure 4. Just 14 per cent of UCL usage related to the student halls of residence network (hor), 44 per cent to users of the WTS Staff, Cluster and Remote Cluster services network (uclusers), 20 per cent to general UCL usage (ucl), 7 per cent to the philosophy (phil) network, 4 per cent to the genetics (gene) network and 2 per cent to the School of Library, Archives and Information Studies (SLAIS).

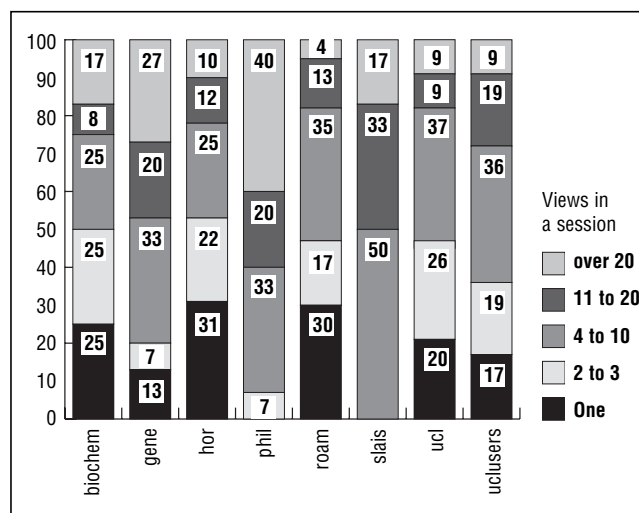
Figure 4:
Percentage distribution of page views by UCL sub-network



Examining the subject of the e-book viewed by sub-network label provides interesting information. OSO e-books are categorised by main subject: religion, political science, philosophy, and economics and finance. Users accessing via biochemistry and gene networks mainly looked at philosophy pages: 100 per cent and 87 per cent did so. This suggests that these networks are used by students from these subjects who study philosophy options and/or that some of the philosophy books appealed to staff on these networks (history of ideas/science etc). Users on the hor (halls of residence) network mainly looked at political science content with 77 per cent doing so. Unsurprisingly, users on the philosophy network mainly viewed philosophy titles (97 per cent), which confirms the accuracy of the sub-network label. About half (47 per cent) of Roam (Roamnet wireless network) users viewed political science content, while a third (32 per cent) viewed philosophy content.

In terms of the number of views in a session, the networks where particularly busy or deep sessions were conducted were: philosophy (60 per cent of sessions viewed 11 or more pages) and SLAIS (50 per cent). Thirty-one per cent of sessions located at the halls of residence and 30 per cent conducted over wireless network recorded just one view in a session, in other words they were particularly light or short sessions. What is particularly noticeable from the sub-network analysis is how different the session profiles of the subject departments were (Figure 5).

Figure 5:
Percentage distribution of number of pages viewed in a session by sub-network



In terms of session length two thirds of SLAIS sessions lasted three minutes or longer. In contrast over half of sessions located in biochemistry (66 per cent), halls of residence (53 per cent) and wireless network (56 per cent) lasted under three minutes.

Sessions undertaken by philosophy saw the print facility employed most: 47 per cent of sessions printed off at least one page.

Lastly, there were considerable differences according to referrer link, with half of all sessions conducted by users from biochemistry (50 per cent) and genetics (50 per cent) accessing the service via Oxford Scholarship Online, while 78 per cent of philosophy sessions emanating from search engine users.

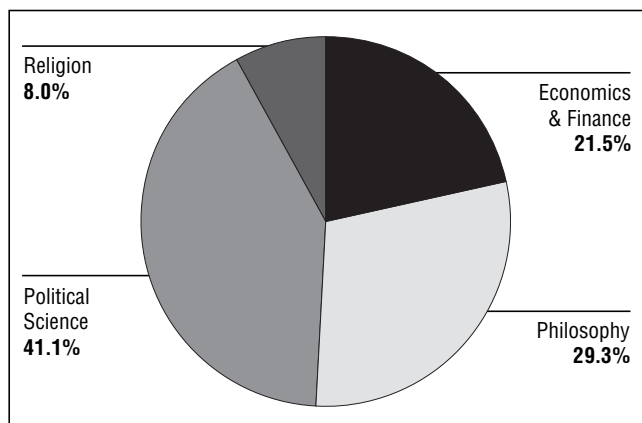
6.2.7 Subject of e-book viewed

Subject use was measured in a number of ways:

- 1) by number of pages viewed;
- 2) by amount of time spent online; and
- 3) by number of e-book titles viewed.

As Figure 6 shows, in terms of OSO's subject classification, 21.5 per cent of views were to economics and finance titles, 29.3 per cent to philosophy titles, 41 per cent to political science and 8 per cent to books on religion. This needs to be read in the context of the subject distribution of the available e-books. Thus 24 per cent were on religion, 19 per cent on economics and finance, 35 per cent on philosophy and 23 per cent on political science. In the light of this, political science use was much greater than might have been expected and that of books on religion much less than might have been expected, although this is far from surprising given that UCL has no department of religious studies.

Figure 6:
Percentage frequency of views by subject of the book



Subject usage varied quite considerably from month to month, adding further to the picture of volatility that is building in regard to e-book use. Thus there were a greater number of page views to political science titles in March. Indeed, well over half of all the views made in March were to political science titles although this subject accounted for just a fifth of views in January and a third in February. Similarly, religious titles made up just 5 per cent of use in February and March, which was considerably down from the 22 per cent recorded in January. In the case of philosophy the relative share of usage fell in March and made up just 17 per cent, as compared to over a third in the previous two months. It is likely that these changes occur as a result of changes in the modules studied.

Other differences worthy of highlighting were:

- Users of political science titles were less likely to view a table of contents with 12 per cent doing so compared to about 15 to 21 per cent for other subjects.
- Almost half of religious book pages were viewed off-site. In direct contrast two thirds (72 per cent) of philosophy titles were viewed on the UCL campus.

Religious titles recorded the longest view time of 19 seconds and Philosophy (12 seconds) the shortest.

Number of titles viewed

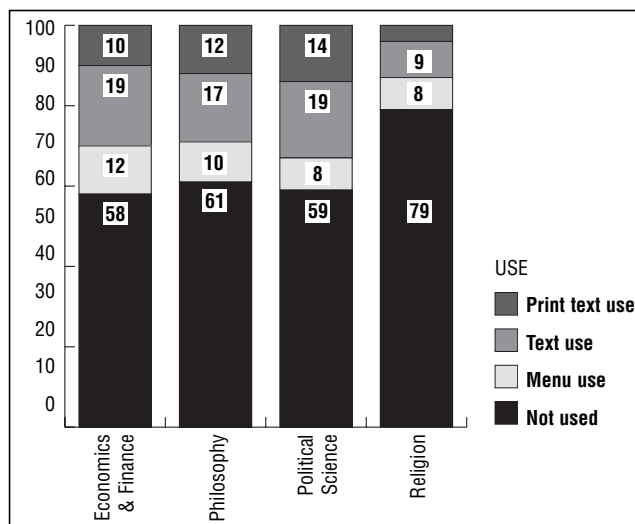
There are two aspects to this usage analysis. First, there is the proportion of books used or not used, a metric (we call this title reach) which says something about the absolute title usage or reach. This metric does not say anything about the quantity of usage. The second aspect is the volume of use expressed as an average over titles, a metric which says something about the quantity of usage.

Figure 7 examines several ways of measuring title reach:

- 1) the percentage of titles where content pages were used;
- 2) the percentage of titles where menus were used (it is possible that after inspecting the menus the user decided the book was not relevant);
- 3) the percentage of titles which had pages printed using the on-site print facility (a relatively strong indicator of satisfied use); and
- 4) the percentage of titles not used at all.

It can be seen that religious books performed relatively poorly by the latter metric: 79 per cent of titles were not used and just 13 per cent of religious book saw actual content viewed. Political science books, in terms of content viewed, were the most used, with a third of titles used in this way (a figure which includes pages viewed and printed). Political science also recorded the highest percentage of titles which saw their pages printed.

Figure 7:
Percentage of e-book titles used by subject (title reach)



6.2.8 Age of e-books viewed

OUP supplied the date that each book was published in print form. The age of each book was estimated by the difference between current period (2007) and the date the book was published in print and five groupings were derived:

- current titles (9 per cent of titles),
- books 1 to 2 years old (25 per cent of titles),
- books 3 to 6 years old (41 per cent of titles),
- books 7 to 11 years old (16 per cent of titles), and
- books over 11 years old (9 per cent of titles).

In general, use reflected the age distribution of books in the collection. Thus: those published over eleven years ago accounted for 11 per cent of page views and 9 per cent of titles available were of this age; 19 per cent of page views were accounted for by books aged between seven and eleven years and 16 per cent of titles were of this age. Those published three to six years ago accounted for 45 per cent of use and 42 per cent of titles were of this age. Titles one to two years old accounted for 17 per cent of usage and 25 per cent of titles were of this age. Those published in the most current year (2006) attracted 8 per cent of usage and represented 9 per cent of books available. A major difference between e-books and e-journals is that with the latter we would expect far higher usage of the current year's publications. For example, usage of OhioLINK e-journals showed that 55 per cent of use of journal articles in genetics, 54 per cent of botany, 50 per cent of probability and statistics, and 49 per cent of cytology journals concerned articles one or two years old (Huntington, Nicholas, Jamali and Huntington 2006).

Title reach decreased with the age of the book. Thus, while under a quarter of current titles were viewed at least once this was true of only a third of titles aged over 11 years. Current books received the most menu-only views and 13 per cent of these titles received menu only views, suggesting a current awareness use. In terms of page views books aged between three and six years received the most number of views, on average 17 views per title. A view time analysis showed that current year books recorded the longest view time of 20 seconds. Books 1 to 2 years old were viewed for about half this length, while older books recorded a view time of 13 to 16 seconds.

There were some interesting differences by subject and physical location of the user:

- Economics and finance recorded the greatest use of books aged over 11 years; 23 per cent of content viewed

was of that age. Philosophy, perhaps surprisingly, made the greatest use of current material and 18 per cent of use was to books published in 2006.

- There was a tendency for UCL site users to view older material; about a third viewed books seven or more years old compared to a quarter in the case of off-site users.

6.2.9 Catalogued books

Thirty six per cent of the OUP e-books were catalogued to establish whether this had an impact on usage. The number of e-books catalogued was 438, with 234(19 per cent) being CURL catalogued and 204 (17 per cent) being UCL catalogued.

We found that cataloguing had a positive impact. A UCL-catalogued book was twice as likely to be used as a non-catalogued one. In addition, 44 per cent of UCL-catalogued books were used compared to 29 per cent of CURL catalogued and 20 per cent of non-catalogued titles. The researchers hypothesise that lecturers are unlikely to recommend readings if the books are not in the UCL library print catalogue. CURL e-book catalogue entries would be for print books not held at UCL and hence are unlikely to appear on recommended reading. Hence there are two effects at work here: a catalogue effect and a, hypothesised, lecturer recommended reading list effect.

Usage per title also varied by catalogue status, with UCL catalogued e-books attracting over twice the page views as compared to non-catalogued and CURL catalogued options: 19 views per title compared to 10 for non-catalogued titles. CURL catalogue books scored a figure midway (15) between UCL and titles not catalogued. The greater relative use of UCL catalogued books, argues that students are relying on catalogued access. In particular what is believed to occur is that students scan the lecturer list of reading and then go to the library catalogue and attempt to locate the reading material.

7 Conclusion

This paper represents a release of log data from the SuperBook Project and the first e-book usage findings based on deep log analysis methods. It presents an analysis of e-books usage on Oxford Scholarship Online, one of three e-book packages introduced to UCL during the autumn/winter of 2006. Two new metrics of usage have emerged from the work:

- a) title reach, that is the proportion of books used or not used; and
- b) the volume of use expressed as an average over the population of e-books.

The main conclusions of this study were:

- Even with limited promotion the collection obtained relatively high levels of use, with nearly 11,000 pages viewed in three months.
- Most of those who found the service viewed a relatively large number of pages. Well over half of all sessions saw more than four pages viewed, and one in nine saw more than ten pages viewed. This represents high levels of viewing compared to e-journals. Furthermore, a good proportion of users were clearly taking advantage of the rich choice of titles available, with one in four sessions seeing more than three books viewed.
- Over two-thirds of usage took place on-site. There were differences between on-site and off-site users. Off-site users adopted a more direct approach, nearly three quarters of views were to full-text pages; the figure for on-site users being just over being half. Furthermore, off-site users were more likely to view an e-book with one in ten sessions not recording a view to a book as compared to four in ten for on-site users. These differences may reflect access expectation on the part of the users. Users within UCL may expect the

service to be maintained hence they may just browse to see what is there with the intention of accessing the books later, at a time when they want to use them.

- Those people accessing via a search engine were most likely to record more views in a session and were more likely to view text pages. Perhaps, in the relatively early days of e-book access, search engine users did not expect to find the material free, in other words they had a negative access expectation, and were eager to grab the opportunity to squirrel away pages.
- Use was highly concentrated with just two of the over 1200 titles available accounting for over 12 per cent of the page views and the top 20 titles accounting for 43 per cent of usage. Something which hints at the future potential of e-books.
- Subject usage varied quite considerably from month to month, adding further to a picture of volatility that is building in regard to e-book use. Well over half of all the views made in March were to political science titles although this subject accounted for just a fifth of views in January and one-third in February. This could be due to the fact that we are dealing with students whose e-book attention span lasts only as long as the module (ten weeks).
- A major difference between e-books and e-journals is that with the latter we would expect a higher usage of the current year's publications. Thus for e-books 45 per cent of use was accounted for by titles published three to six years ago. However current books did receive the most menu-only views, suggesting a current awareness use perhaps.
- Catalogued books were much more likely to be used, in the case of UCL catalogued e-books they attracted over twice the usage as compared to non-catalogued books. The researchers hypothesise that lecturers are unlikely to recommend readings if the books are not in the UCL library print catalogue.

References

- Armstrong C, Edwards L and Lonsdale R (2002) 'Virtually there? E-books in UK academic libraries', *Program: Electronic Library and Information Systems*, 36 (4), 216-27.
- Chu, Heting (2003) 'Electronic books: viewpoints from users and potential users', *Library Hi Tech*, 21 (3), 340-6.
- Huntington P, Nicholas D, Jamali H R and Tenopir C (2006) 'Article decay in the digital environment: a usage analysis by date of publication employing deep log methods', *Journal of the American Society for Information Science and Technology*, 57 (13), 1840-51.
- Langston M (2003) 'The California State University e-book pilot project: implications for cooperative collection development', *Library Collections, Acquisitions, & Technical Services*, 27 (1), 19-32.
- Levine-Clark, M. (2006) 'Electronic book usage: a survey at the University of Denver', *portal: Libraries and the Academy*, 6 (3), 285-99.
- Nicholas D, Huntington P, Jamali H R and Tenopir C (2006) 'What deep log analysis tells us about the impact of big deal, Case study OhioLink', *Journal of Documentation*, 62 (4), 482-508.

Contact

David Nicholas
david.nicholas@ucl.ac.uk