POSTPRINT
Published Journal of Academic Librarianship 33(4), 446-453, 2007.

The use, users and role of abstracts in the digital scholarly environment

David Nicholas
Director of CIBER (Centre for Information Behaviour and the Evaluation of Research) and Professor at and Director of the School of Library, Archive and Information Studies (SLAIS), University College London.

Address: School of Library, Archive and Information Studies (SLAIS), University College London, Henry Morley Building, Gower Street, London WC1E 6BT
Tel: +44 20 7679 2107
Fax: +44 20 7383 0557
Email: david.nicholas@ucl.ac.uk

Paul Huntington
Senior Researcher and a founder member of CIBER, SLAIS, University College London

Address: School of Library, Archive and Information Studies (SLAIS), University College London, Henry Morley Building, Gower Street, London WC1E 6BT
London WC1E 6BT
Tel: +44 20 7679 7205
Email: p.huntington@ucl.ac.uk

Hamid R. Jamali
A researcher in CIBER, SLAIS, University College London

Address: School of Library, Archive and Information Studies (SLAIS), University College London, Henry Morley Building, Gower Street, London WC1E 6BT
Tel: +44 20 7679 7205
Email: h.jamali@gmail.com
(The point of contact for all correspondences)

ABSTRACT

Utilising transactional log data taken from digital journal libraries and attitudinal and demographic data derived from a questionnaire survey, the article pieces together evidence concerning the use, users and role of abstracts in a digital environment. It shows that abstracts are used in large quantities, even when full-text viewing facilities are available. The 'popularity' of abstracts is partly a function of how users navigate

towards content in cyberspace 'through search engines and gateways' and partly because they provide a quick and effective means of assessing relevance of content.

INTRODUCTION

One of the most interesting, but possibly unexpected findings of the Virtual Scholar Research programme[i] is how popular abstracts proved to be. [1] In a digital world, rich in full text documents, abstracts remain ever-popular, even in cases where the user has complete full-text access. In order to find out more about the popularity and role of the abstract in the digital scholarly environment we have brought together data on the use and users of abstracts extracted from three log studies of major digital journal libraries - ScienceDirect, OhioLINK, and Oxford Journals. This evidence-base covering hundreds and thousands of online views, made by tens of thousands of users, has been complemented by attitudinal and demographic data from two questionnaire studies, one that was sent to 750 ScienceDirect users and the other sent to users of the Oxford University published journal, *Nucleic Acids Research*, in which over a 1000 people answered questions about their use of abstracts.

Although the size of the total population covered is indeed a very large one, clearly we need to be cautious in regard to interpreting the data as we have pooled data together from different studies, conducted at different times (over a period of two years). It is also worth pointing out that the population surveyed has a bias towards science. A unifying and important characteristic of all these studies is that all employed a common methodology, deep log analysis.

[i] http://www.ucl.ac.uk/slais/research/ciber/

LITERATURE REVIEW

Abstracts, despite having a long history, are still the subject of much comment, debate and research. Pinto and Lancaster[2] pointed out that, despite the changing role of abstracts from 'a human-readable output from electronic databases', to 'computer-searchable surrogates for larger bodies of text' and with the arrival of ubiquitous full-text 'They remain useful summaries to be read by humans' and protect the searcher from 'unacceptable levels of irrelevancy' that result from full-text searching. New forms of abstract are seemingly always being dreamed up (structured abstracts being the latest) and merit much comment and evaluation.[3] The quality issues seem always to be with us, too.[4] Pinto [5] in probably the most relevant paper found considers abstracts in the digital environment and suggests that the traditional abstract is not up to the job and proposed some changes. Despite the fact that abstracts continue to be a subject of comment that is not to say we always learn anything new about them. Thus we are still told that students do not use them,[6] although the evidence presented is weak and it is something we shall later dispute.

Much of the existing literature on abstracts addresses issues concerning information retrieval[7], automated abstracts [8] and different formats [9]. However, what appears almost wholly lacking from the recent published literature, certainly in respect to the digital environment, are user studies, studies which tell us who is using them and why? And log studies provide much I this regard as this paper will demonstrate.

METHODOLOGY

As mentioned earlier the main means of obtaining abstract use and user data was by through deep log analysis, a methodology pioneered by the authors and featured in an earlier article in this journal[10]. Digital information platforms have a facility by which logs are generated that provides an automatic and real-time record of use by anyone who uses them. They represent the digital information fingerprints of the users and by analysing them information seeking behavior can be tracked and mapped, in our case regarding abstract use. When these fingerprints are enhanced - we call this deep log analysis (DLA), they tell us about the kinds of people that use the services. DLA works with the raw log data, not the processed, pre-defined and selective data normally provided to the publisher and library participants. That way more accurate, detailed, and panoramic pictures of digital information seeking behavior can be produced. However, logs do not record all user transactions (caching is a problem), user session information can become muddled (proxy-servers are the problem) and because of dynamic IP addresses users are difficult to define (for more information on problems regarding log analysis see, Jamali et al.[11]). Hence, logs can only offer an indication of likely underlying behavioral patterns that need to be followed up by follow-up surveys and more qualitative methods, which is exactly what we have done in this paper. For more details of the methodology see Nicholas et al.[12]

The logs referred to in this paper were standard logs and detailed the internet protocol (IP) address, date, time and the page viewed by a user. The following OhioLINK log is a typical example of the kind of logs we analysed to obtain data about abstracts.

```
129.22.7.22 - - [31/May/2004:00:06:39 -0400] "GET /cgi-bin/sciserv.pl?collection=journals&journal=02663538&issue=v61i0006&article=889_tdpogerdp&form=pdf&file=file.pdf HTTP/1.1" 302 5 "Mozilla/4.0 (compatible; MSIE 6.0; Windows 98; .NET CLR 1.1.4322)" "-"
```

The first field (129.22.7.22) details the user's internet protocol (IP) number. Users are not required to enter a user name or password to enter OhioLINK. OhioLINK authenticates entry based on the user's specified fixed IP addresses or range of IP addresses. The second field [31/May/2004:00:06:39 -0400] details the date and time of the transaction and how many hours off Greenwich meantime this is. The third field details the item delivered in response to a request from the user: "GET /cgi-bin/sciserv.pl?collection=journals&journal=02663538&issue=v61i0006&article=889_tdpogerdp&form=pdf&file=file.pdf. The number 02663538 is the ISSN number of the journal, although the hyphen after the fourth digit is missing. The number v61i0006, in the logs, specifies the volume and issue number in this case the number relates to volume 61 issue 6. The alpha-numeric sequence 889_tdpogerdp identifies the article. Lastly "file.pdf" tells us the file was supplied in a PDF format, in fact almost all article documents were supplied as PDF. The fourth field details the status of the download. The delivery status field records whether the page was delivered without any problems. The fifth field records the browser details and is the record of the browser that the client is using to attach to the Internet.

Special care needs to be taken when measuring and interpreting abstract use recorded in the logs because:

1) Abstracts are typically free to view for all users, not just to subscribers of the digital service as is generally the case with full-text articles. Therefore, a proportion of users, coming in from search engines, for instance, could only view an abstract – they had no choice, it was a substitute for the real thing. This was the case with ScienceDirect, where the majority of users studied were subscribers. In the case of *Nucleic Acids Research* (NAR) for most of the period covered. About half of NAR users were non-subscribers. However, it was different in the case of the OhioLINK Electronic Journal Centre, which is a closed system, with only subscribers being allowed to search the system and full-text available to all - there was a genuine level playing field here.
2) In the case of some digital libraries abstracts have to be viewed before you can see a full-text item – and this boosted their viewing figures of course. This was the case with the ScienceDirect but not with the OhioLINK and NAR studies, where users were given a choice as to what they viewed.

THE PROJECTS AND THEIR FINDINGS

Authors as users: a deep log analysis of ScienceDirect[ii].

Background

In this study, conducted during the period 2005-6, a questionnaire supplying user demographic data and attitudes towards scholarly communication activities of 750 authors were directly related to their usage (logs) of ScienceDirect for an eighteen month period. The online questionnaire went out on May 2005 to 49,266 scholars. The mailing list was largely formed from a database of scholars Elsevier regularly use to obtain feed back from scholars on authorship and scholarly related matters. Most of

[ii] http://www.sciencedirect.com

these were Elsevier authors, many of whom had published in ScienceDirect. The respondent's IP number was collected as well as their written responses. This database of anonymous IP numbers was then compared to IP numbers collected by the server used to access the ScienceDirect web site over the previous 18 months. Those IP numbers that could be safely associated with a single computer were selected. Thus, in order to derive at a unique user, we only extracted data from the transaction log file for those IP addresses that had a unique machine cookie. Machine cookies are dropped on a client computer when a user accesses ScienceDirect. If an IP address in the log file has a number of different machine cookies associated with it then one of the following might be true: more than one user is using the machine and this would cover proxy IP addresses; the client machine is allocated different IP numbers on different occasions as would be the case with those client machines that are part of a floating IP network; or lastly that the clients machine is deleting cookies either automatically or by the user. For this study only log entries for IP addresses with one machine cookie were used. That is most of the anomalies surrounding non-unique IP addresses as a result of proxy servers and floating IP addresses have been excluded and hence a pretty good mapping of individual user, who filled in the questionnaire, to log entries was obtained.

During the survey period authors conducted 17,000 sessions, which saw 110,000 pages being viewed. A little under a third of users (30%) accessed the site via a gateway site - that is a site that provides a link to a ScienceDirect full text item. This link may be a hyperlink in an article or abstract residing on another site. Where this is an abstract link the abstract may be the abstract of the linking article but the abstract is

held on the external site as in the case of PubMed and PubMed Central[iii]. Hence not all abstract views will be recorded in the logs, so what we are talking about are minimum figures.

Type of article item viewed

Session analyses are particularly illuminating in understanding abstract use as they demonstrate the combination of items that are viewed as a result of an online search session. Thus a distinction can be made between sessions where just an abstract was viewed, sessions that just viewed a PDF or a full text (HTML) item or a session in which a combination of items were viewed. This was undertaken in the case of the ScienceDirect study and this showed that most (33%) sessions saw a PDF item being viewed; 20% saw an HTML version and a PDF viewed; the same proportion (20%) just viewed an abstract; 14% just viewed a full text, and 13% viewed either an abstract and an HTML or a PDF. Fig. 1 relates. Thus over a third of sessions saw an abstract being viewed, and given the aforementioned gateway use the true figure is undoubtedly higher than that.

Take in Figure 1 Percentage breakdown of type of article item viewed (ScienceDirect).

Characteristics of abstract users

The ScienceDirect study proved particularly fruitful here as it was possible to relate the digital fingerprints of users, as recorded in the logs, with demographic data they provided in a questionnaire study. However, given the sample size these data should

[iii] http://www.pubmedcentral.nih.gov

be interpreted with care, although the number of sessions covered was impressive. In terms of subject (Fig. 2) Social Science users, perhaps surprisingly, conducted the highest proportion of sessions where just abstracts items were viewed - 41% did so. Mathematicians were close with 40% of their sessions seeing only abstracts viewed (40%); Computer Scientists (35%) and Economists (33%) followed some way behind. By contrast Chemists (12%), Engineers (13%) and Life Scientists (13%) conducted relatively few abstract only sessions. Indeed, in the case of chemists, even if we include sessions where full-text and abstracts were viewed then only about one fifth (21%) of sessions viewed abstracts; the equivalent figure for social scientists was more than half (51%). Of course, he Life Sciences figures were depressed by PubMed use.

Take in Figure 2 percentage breakdown of type of article item viewed by subject (ScienceDirect).

Fig. 3 shows that abstracts were more popular with academics (36% of sessions saw an abstract viewed) and least popular with users from research institutes (20% featured abstracts). Medical Centres that were part of universities were included in the Medical Centre count.

Take in Figure 3 Percentage breakdown of article views by format and organisational affiliation (ScienceDirect).

With the exception of the under 36s, the number of abstract only sessions conducted, tended to increase with age, (Fig. 4). About 14% of those 36 to 45 just undertook an

abstract item session but this increased to 29% for those aged 56 to 65 and to 35% of those 65 and over. This could be a result of the time demands on senior people or it could be because, with time, they have learnt to determine relevance this way. If we factor in to the analysis mixed sessions, where abstract and full text were viewed then the differences between the 56-65 and the 65+ groups shrinks a little as those over 65 conduct relatively few mixed sessions.

Take in Figure 4 Percentage breakdown of article views by format and user's age (ScienceDirect).

Significant differences emerged in regard to an analysis of (self-reported) occupational status (Fig. 5). Thus junior researchers were much more likely to conduct abstract only sessions (25% were of this type), than senior staff (18%). There was no sign that (Postgraduate) students were avoiding abstracts but there was further evidence of students' appetite for full-text/html. Another CIBER study of an e-journal database (Blackwell Synergy)[13], which broke down abstract use by types of student (undergraduate and postgraduate) in fact showed that undergraduates were the biggest users of abstracts (by a 5% margin over all other groups) and these were all subscribers so it can not be simply put down to poorer access to full-text.

Take in Figure 5 Percentage breakdown of article view by format and user's occupational status (ScienceDirect).

Users from Australasia were surprisingly very heavy users of abstracts, approaching three quarters (71%) of their sessions viewed abstracts (Fig. 6). Eastern Europeans were also proportionately big conductors of abstract only sessions (47%). By contrast

South Americans appeared to have an aversion to abstracts (only 14% of sessions saw an abstract viewed). The Synergy study, using more specific geographical units confirmed relatively low levels of abstract use for South America (Brazil in fact)[14]

Take in Figure 6 Percentage breakdown of article views by 9 world regional groupings (ScienceDirect).

Determining the impact of open access publishing on use and users: case study Nucleic Acids Research (NAR)[iv].

Background

Two studies were undertaken during the period 2005-2006: a) a log analysis investigation covering the period January 2003 to June 2005, to determine the impact of open access publishing on use; during which time 1.5 million separate internet protocol numbers (users) were identified to have made about seven and half million search sessions and to have viewed approximately thirteen and half million unique items (abstracts and articles); b) a questionnaire study of *NAR* users, which sought general feedback in regard to the move of the journal to an open access model and general information on scholarly information seeking behavior. This survey included questions about their use of abstracts. 13,396 emails were sent out in 2006, containing a link to online survey, to *NAR'*s online submission/peer review database and 1145 respondents completed the survey (~9% response rate). The population of NAR is fairly homogenous and this is no reason to doubt the representativeness of the sample.

[iv] http://nar.oxfordjournals.org

Again, abstract views were under-recorded because users coming in from PubMed[v] would have viewed the abstract there and only clicked through to the *NAR* site for the full-text.

Type of article item viewed

The findings for *NAR* were similar to those for ScienceDirect but with slightly higher views of the full-text recorded - not surprisingly given the greater freedom to view full-text. Thus article only sessions accounted for about two-thirds of use (73% in the case of ScienceDirect), abstract only sessions accounted for a quarter and sessions which saw both viewed accounted for approximately 10% of sessions. This meant that 35% of sessions saw an abstract viewed (the ScienceDirect figure was 33%). Given that users had a choice whether to view an article or abstract (i.e. the user did not have to view an abstract before an article was viewed) the fact that one in four sessions just viewed abstracts is evidence of the popularity of the abstract.

Importantly, the *NAR* study also provided an opportunity to discover what actually happened to abstract use once all subscription controls on the viewing of full-text articles were released as a result of the journal moving to an open access (OA) model. Table 1 shows the before OA and after OA article and abstract viewing figures. Not surprisingly, as result of moving to an OA model, the early signs were of a shift from abstract only sessions to a greater use of articles. However, the initial growth in usage first occurred in the case of mixed sessions, where articles and abstracts were viewed (a 25% increase in January); article only sessions went up less strongly (a 19% increase in January). Since February though, article only sessions have demonstrated a more robust growth trend.

[v] http:// www.ncbi.nlm.nih.gov/entrez/

Take in Table 1 Session types by item viewed - daily average figures (5% trimmed mean) for November 2004 to March 2005 (NAR).

The views and perceptions of scholars

Users of *NAR* were sent a questionnaire to find out what they thought of their journal having moved to an open access model. In addition they were asked questions about their general information seeking behavior and among these questions were a few about their use of abstracts, prompted by the strong abstract usage that had been witnessed in the logs. There were three closed questions (on viewing behavior, citing habits and relevance checking) and one open ended question which asked them to reflect on their use of abstracts in the case of *NAR*.

Views were sought on what article items were viewed in order to see whether what users said bore out what we had discovered in the logs. Thus the following question was asked:

Q. An *NAR* article of potential interest to you is available online and you are given the following options:

- View Abstract only
- View Full-text HTML
- View Full-text PDF

Q. Which option do you usually take first?

- Abstract.
- Full-text HTML
- Full-text PDF
- It varies too much to say:

The number of respondents to this question was 1145, of which 37% said they viewed an abstract first, just 11% said HTML, 46% said PDF and 7% said it varied

too much to say. The results point to the popularity of abstracts and, indeed, the abstract only figure was higher than the abstract only figure for NAR (25%), but if we add in mixed sessions (there might have been some confusion in respondent's minds) where both were viewed (35%) then the figures are not that far apart.

It is sometimes said in discussion, but rarely proven, that today's virtual scholars are so busy that they often only have time to read the abstract and this line of questioning provided another opportunity to investigate the extent to whether users relied wholly on abstracts. The question was framed as following:

Q. Do you always read the full paper before you cite it in your work?

- Yes:
- No:
- Depends:

1145 people responded to the question and the surprise has to be that barely little more than half (52%) said they always did. 40% said it depended and a further 9% said no.

Abstracts are said to help out with relevance assessment and the following question sought user's views on this:

Q. For approximately what percentage of time can you determine from the abstract whether an *NAR* article will be of interest to you?

- 0-20%
- 20-40%
- 40-60%
- 60-80%
- 80-100%

Clearly abstracts were used in this way, successfully so, with nearly one third (32%) of the 1144 respondents saying they could determine whether they were

interested in an abstract 80% or more times and this figure rose to 71% when we include those respondents saying 60-80% of the time. The figures for the rest of the percentage bands were: 0-20%: 5%; 20-40%: 8%; 40-60%: 16%.

An open ended question sought to take our enquiry further: 'Our logs show considerable numbers of views to the abstracts of *NAR* articles. On the basis of your own behaviour can you tell us why this should be so? The majority of the 929 people who responded mentioned that they viewed abstracts to determine whether the article was of interest to them, to see what is contained in terms of content, methods etc. As one respondent explained: *"When I scan the literature, I do a very quick screen on the titles first. For the rare articles (from many journals) that make that cut, I then read through the abstracts to see if I wish to read more. I then pull down the full article only for a select few."*

For some, abstracts were a substitute for the article itself. Respondents said that they would read just the abstracts and get the information they needed without having to read the full-text of articles in some cases. About 3% mentioned this and the following quote is illustrative: *"In some cases the abstract even provides me enough information and I don't need to read the full paper. Given the amount of papers published, good scientists only rarely can afford reading a full paper."*

Interestingly, the previously mentioned ScienceDirect log study showed that the user was more likely to substitute if the article was a long one.

Time was another factor mentioned by respondents; however, it was only mentioned by 8% of respondents. Abstracts were used to determine whether the article was worth

the effort (and time) to download its PDF. About 3% mentioned that that they use abstracts because they were faster to access and download. "*I usually scan an abstract for points of interest as this is faster than waiting for the entire PDF to load. In many cases I can tell from the abstract that I do not need to look at the full paper. Loading of the abstract takes a second, of the PDF 5-10 seconds."* There were some exceptions and those people with fast internet they said they tend to look at PDF anyway: *"I often read the abstract on the screen as a PDF file, but it is more logical often to read the abstract first and then decide whether to download the PDF. This is particularly true when you have a slow internet connection. I have a fast one and often download the PDF first."* They were also preferred by users operating in interdisciplinary fields who had to cover a lot of ground. "*I am from an interdisciplinary field I have to scan approx 50 journals a month. This would not be possible by trying to get access to the full articles."*

Finally, for some people, abstracts were all they were entitled to view.

Not everyone liked abstracts. Thus there was a small minority (1.7%) of respondents said that they usually preferred to read the full-texts instead of abstracts. Some people downloaded the full-text even if they were only going to read the accompanying abstract. Those users who had a fast internet connection tended to belong to this group:

Some respondents went on to speculate about largely technical reasons for high use of *NAR* abstracts:

a. When users link in from other sites that normally leads to an abstract and not full-text;
b. Users following 'cited by' links or following references were normally led to an abstract.
c. Links from Google or Google Scholar typically link to abstracts.

MaxData: OhioLINK[vi] study. 2005-7. An ongoing study funded by US Institute of Museum and Library Services. Log analysis was conducted for the period June 2004 to December 2004. During this period 1,215,000 items were viewed on campus in October 2004 and 1,894,000 items viewed off campus between June and December 2004. All users were subscribers with complete full-text rights. Further details of the study can be found in Nicholas et al.[15]

Type of article item viewed

OhioLINK Users could obtain content in the form of an abstract, an article (full-text) or both an abstract and article. Just 9% of views were just to an abstract, 69% were to an article and 22% of titles were viewed in both an article and abstract format. The lower abstract only views can be explained by the complete freedom of choice afforded to all users to view an abstract or full-text article. Nevertheless 31% of sessions saw an abstract viewed, which was not far removed from the figures for ScienceDirect (33%) and NAR (35%).

As was the case with the ScienceDirect study it was possible to examine abstract viewing by subject. Figure 7 show the proportional use of abstracts and articles used by subject of journal viewed and it shows clearly that Physics and the Arts and

[vi] http://www.ohiolink.edu

Humanities make greatest use of abstracts. In contrast with the ScienceDirect study, where the classification of some subjects was somewhat different (the arts and humanities did not feature) and the form of analysis a little different, the big users of abstracts were the Social Sciences and Mathematics, although Physics was up there too.

Take in Figure 7 The proportion distribution by subject articles and abstracts.

We have founded in some of our other studies that the way people navigate to content has a big impact on what they view[16]. Thus search engine users typically viewed more items, subjects and journals in a session and we were interested in determining whether there was an impact on abstract viewing. OhioLINK users could navigate to content in a number of different ways, via alphabetic or subject lists or via a search engine or via a mix of these approaches. Figure 8 gives the distribution of whether a user viewed an abstract only, an article only or both an abstract or an article during a session by type of access. Those using an abstract must have regarded this format as being sufficient to meet their needs; equal to or better than an article, because they clearly could have viewed an article for the same amount of effort. It is clear from Figure 8 that those using the search engine were more likely to view an abstract as well as an article in their session: 54% did so compared to 24% of those that used a subject list search and 15% of those that used an alphabetic journal list search. The explanation for these real differences lies, perhaps, in the fact that search engine users have a greater number of items to view and they resort to abstracts to make a quick selection – abstracts are plainly quicker to load and read. There is another possible reason and that is search engine users by definition are in a hurry or have less need for detail.

Take in Figure 8 Distribution of whether a user viewed an abstract only, an article only or an abstract or an article in a session by method of navigation.

CONCLUSION

The logs and the respondent's comments all point to the fact that abstracts are very popular and, as noted previously, if anything our figures under-play their popularity and usefulness. The popularity of abstracts shown in the logs is partly a function of how users navigate towards content in cyberspace – search engines and gateways tend to point them to abstracts in the first instance. And, of course, those without full-text rights – the non-subscriber or the 'disenfranchised', have no choice at all but to view an abstract and then possibly obtain the full-text from a library. Abstract use did vary from user to user and were particularly popular with older users, social scientists and Australians.

It is clear that abstracts have a key role in the information seeking behavior of scholars and are important in helping them deal with the digital flood. We have demonstrated elsewhere that digital scholars adopt a bouncing or flicking kind of behaviour to help them stride across the digital information universe[17]. Scholars adopt a horizontal rather than vertical form of information seeking which results in users seldom delving deeply into a website and often not returning to a website. It is not difficult to see how abstracts are very suited to this form of behaviour and possibly help encourage it.

Finally, it should be pointed out that the article has drawn together material from a number of studies, employing slightly different methods and conducted over a period

of two years, when considerable change was going on. And in this regard the data should be treated with caution. Nevertheless, we believe the studies we source are complementary and that they all point to the same finding, that abstracts are popular and increasingly used as a means of overcoming the problems of identifying relevant documents from a very large information universe. Whilst conducting the NAR study one of the publishers joked, that given the evidence showing the key role of abstracts in today's crowded information environment, maybe they should reverse their business model and give full-text away and charge for abstracts. They were only half-joking.

REFERENCES

[1] David Nicholas, Paul Huntington, Hamid R. Jamali & Anthony Watkinson, "The information seeking behaviour of the users of digital scholarly journals," *Information Processing & Management* 42 (5) (2006):1345-1365.

[2] Maria Pinto & F.M. Lancaster, "Abstracts and abstracting in knowledge discovery," *Library Trends 48* (1) (1999): 234-248.

[3] Alan J. O'Rourke, "Structured abstracts in information retrieval from biomedical databases: A literature survey," *Health Informatics Journal 3* (1) (1997): 17-20; James Hartley (2002), "Do structured abstracts take more space and does it matter?" *Journal of Information Science 28* (5): 417-422.

[4] Maria Pinto & F.M. Lancaster, "Abstracts and abstracting in knowledge discovery,"; A. Wheatley & C. J Armstrong, "Metadata, recall, and abstracts: Can abstracts ever be reliable indicators of document value?," *Aslib Proceedings 49* (8) (1997): 206-213.

[5] Maria Pinto, "Engineering the production of meta-information: the abstracting concern," *Journal of Information Science* 29 (5) (2003): 405–417.

[6] Judith Keene, "From abstract to relevant: Encouraging access to journals," *Library and Information Research News* 28 (89) (2004): 2-12.

[7] See for example: Andrew Booth and Alan J. O'Rouke, "The value of structured abstracts in information retrieval from MEDLINE", *Health libraries review* 14 (3): 157-166; A. Wheatley and C.J. Armstrong, "Metadata, recall, and abstracts: can abstracts ever be reliable indicators of document value?" *Aslib Proceedings* 49 (8) (1997): 206-213.

[8] See for instance: Ronald Brandow, Karl Mitze and Lisa F. Rau, "Automatic condensation of electronic publications by sentence selection", *Information Processing & Management* 31 (5) (1995): 675-685.

[9] See for instance: Takeo Nakayama, Nobuko Hirai, Shigeaki Yamazaki and Mariko Naito, "Adoption of structured abstracts by general medical journals and format for a structured abstract", *Journal of the Medical Library Association* 93 (2) (2005): 237-242.

[10] David Nicholas, Paul Huntington, Hamid R. Jamali & Carol Tenopir, "Finding information in (very large) digital libraries: a deep log approach to determining differences in use according to method of access," *Journal of Academic Librarianship* 32 (2) (2006): 119-126.

[11] Hamid R. Jamali, David Nicholas & Paul Huntington, " Use and users of scholarly e-journals: a review of log analysis studies," *Aslib Proceedings* 57 (6) (2005): 554-571.

[12] David Nicholas, Paul Huntington, Nat Lievesley and A. Wasti, "Evaluating consumer Web site logs: case study The Times/Sunday Times Web site," *Journal of Information Science* 26 (6) (2000): 399-411.; David Nicholas, Paul Huntington, Hamid R. Jamali & Anthony Watkinson, "The information seeking behaviour of the users of digital scholarly journals," *Information Processing & Management* 42 (5) (2006): 1345-1365.

[13] David Nicholas, Paul Huntington, & Anthony Watkinson, "Scholarly journal usage: the results of deep log analysis," *Journal of Documentation* 61 (2) (2005): 248–280.

[14] David Nicholas, Paul Huntington, & Anthony Watkinson, "Scholarly journal usage: the results of deep log analysis," *Journal of Documentation* 61 (2) (2005): 248–280.

[15] David Nicholas, Paul Huntington, Hamid R. Jamali & Carol Tenopir, "What deep log analysis tells us about the impact of big deal, Case study OhioLink," J*ournal of Documentation* 62 (4) (2006): 482-508.

[16]David Nicholas, Paul Huntington, Hamid R. Jamali & Carol Tenopir, "What Deep Log Analysis Tells Us About The Impact Of Big Deal, Case Study Ohiolink," *Journal Of Documentation* 62 (4) (2006): 482-508.

[17] David Nicholas, Paul Huntington, Hamid R. Jamali and Anthony Watkinson "Characterising and evaluating information seeking behaviour in a digital environment: spotlight on the 'bouncer'," *Information Processing & Management*, 43, (2007) in press, DOI: 10.1016/j.ipm.2006.08.007

FIGURES AND TABLES

Figure 1 Percentage breakdown of type of article item viewed (ScienceDirect).

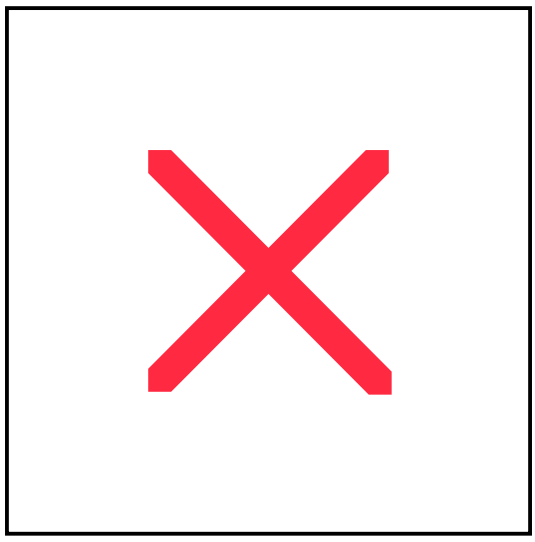


Figure 2 percentage breakdown of type of article item viewed by subject (ScienceDirect).

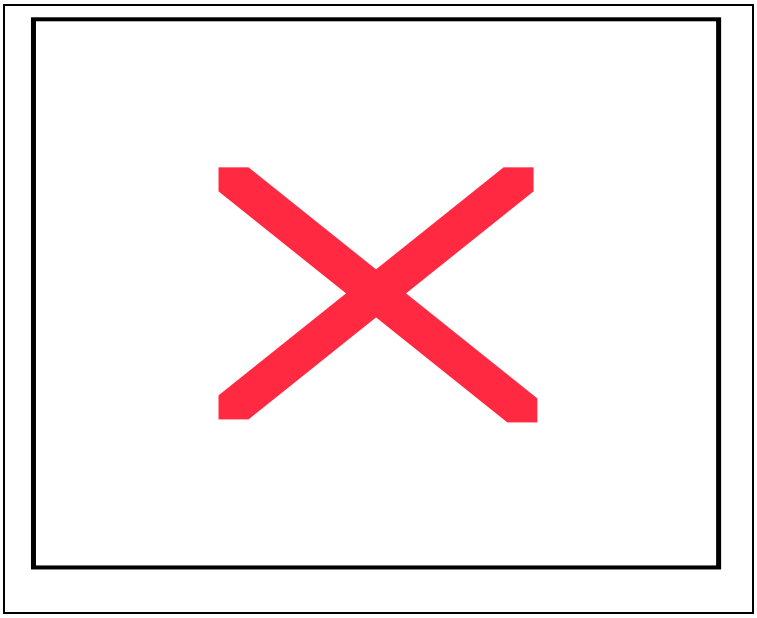


Figure 3 Percentage breakdown of article views by format and organisational affiliation (ScienceDirect).

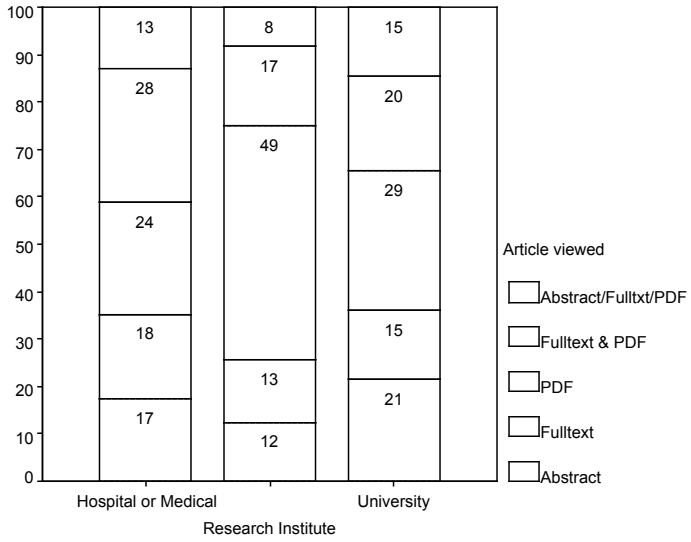


Figure 4 Percentage breakdown of article views by format and user's age (ScienceDirect).

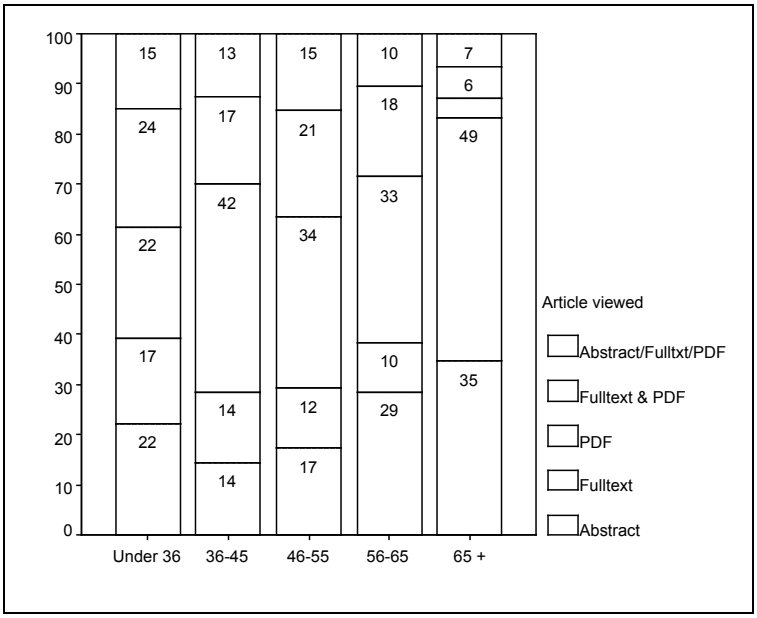


Figure 5 Percentage breakdown of article view by format and user's occupational status (ScienceDirect).

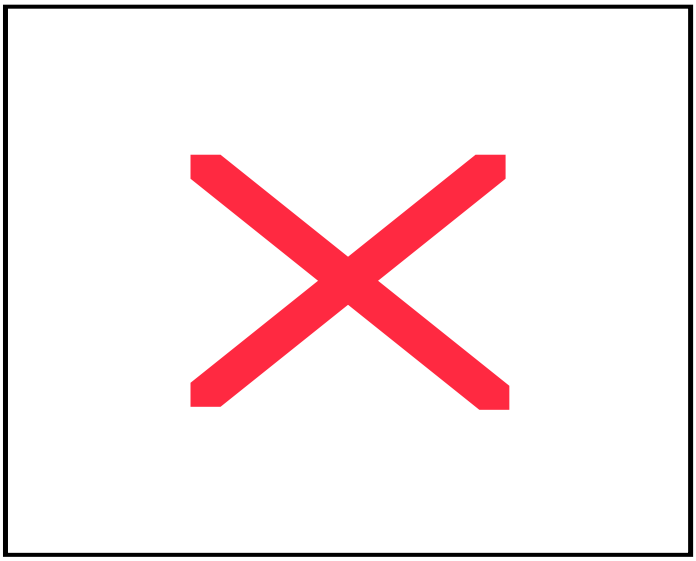
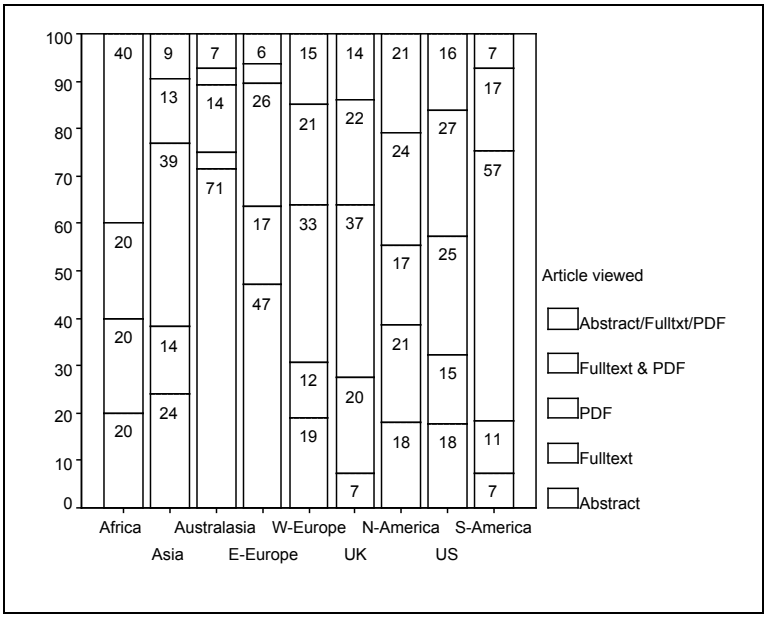


Figure 6 Percentage breakdown of article views by 9 world regional groupings (ScienceDirect).



NB. These are Elsevier geographical groupings and N America refers to Canada and Alaska in this context.

Figure 7 The proportion distribution by subject articles and abstracts.

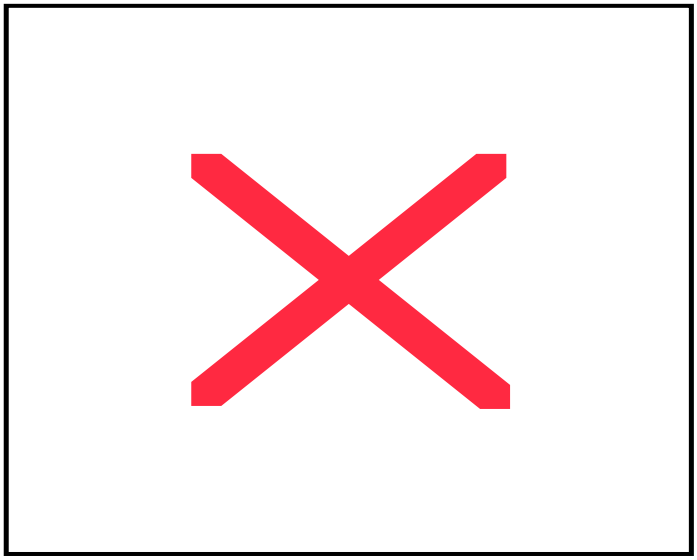


Figure 8 Distribution of whether a user viewed an abstract only, an article only or an abstract or an article in a session by method of navigation.

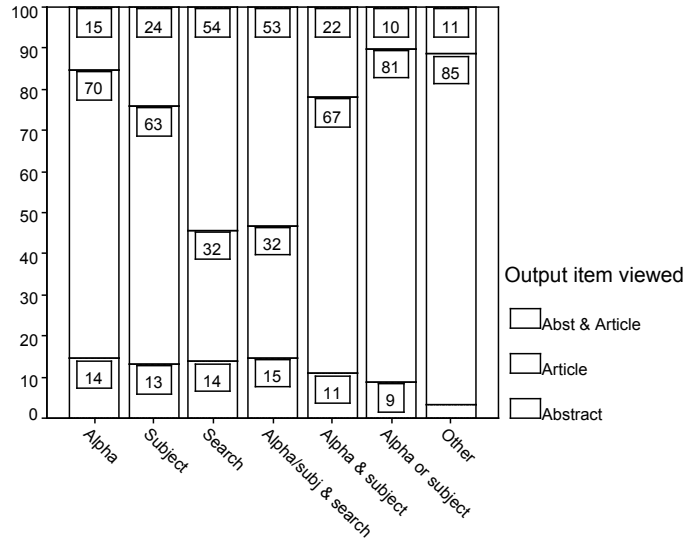


Table 1 Session types by item viewed - daily average figures (5% trimmed mean) for November 2004 to March 2005 (NAR).

Date		Nov	Dec	Jan	Feb	March
Abstracts	2004/05	1,199	1,641	1,909	1,823	1,679
% change on previous month	2004/05	-7.8%	26.9%	14.0%	-4.7%	-8.6%
Article	2004/05	8,081	7,002	8,678	9,450	9,521
% change on previous month	2004/05	3.7%	-15.4%	19.3%	8.2%	7.5%
Mixed	2004/05	2,129	2,403	3,207	2,857	2,989
% change on previous month	2004/05	7.5%	11.4%	25.1%	-12.3%	4.4%