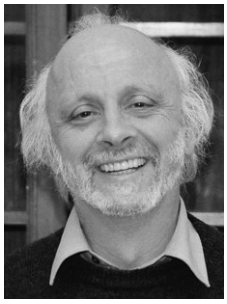
The information-seeking behaviour of the virtual scholar: from use to users

Based on paper presented at the 31st UKSG Conference, Torquay, April 2008

The case is presented for moving on from monitoring activity in the virtual scholarly space to studying the virtual information seeking of users and then relating that to diversity, satisfaction and scholarly outcomes. The article shows that, thanks to new methodological techniques that enable us to obtain deep and robust insights of what goes on in the virtual environment, it is now possible to obtain such data. These methods, using deep log analysis, are outlined and results illustrated in regard to CIBER's Virtual Scholar Research Programme and their use in three major, national research studies – the Joint Information Systems Committee (JISC)-funded 'National E-books Observatory Project', The JISC/British Library (BL)-funded Google Generation study and the Research Information Network (RIN)-funded 'Evaluating the Usage and Impact of E-journals in the UK'.



DAVID NICHOLAS
CIBER

This paper is about users rather than use and that is a very important distinction to make. It is very important because we need to grow up in a digital information sense and seek a better understanding of information-seeking behaviour, and not simply burrow ever deeper into information activity, usage factors and all the rest. Once we have established what has happened to information-seeking behaviour in a digital environment then we can establish what this has led to in terms of scholarly outcomes and what actually represents good information-seeking practice (something that we have shied away from). This would enable us to determine whether in fact the huge and ever increasing investment in digital information resource is bearing fruit. My thesis is that many of us are working with an information-seeking model that is pre-digital and is based upon research into a few hundred people, in a minority of unrepresentative subject fields. What I am hoping to do is to demonstrate how far we all have come in terms of our information-seeking behaviour and how it differs from the models that are often used to run our information services.

There is a real danger of the information profession disconnecting from the user as a result of the profession's pre-occupation with monitoring activity, and not users. Of course, information on the latter is largely absent from the COUNTER-compliant logs. This is particularly worrying because the virtual audience, a majority audience in many cases, differs in composition and size from the traditional (largely hard-copy) library audience. It is also anonymous – we cannot see what is going on in the virtual space unless we employ specialized techniques, like deep log analysis. Furthermore, worryingly for librarians and pleasingly for publishers, the information seeking of the scholar now takes place in publisher cyberspace. While librarians are allowed a peek into this space, publishers now know more about scholars than librarians do. How things have changed.

As a profession we are still too preoccupied with resources and content. Thus at a top-end conference in Fiesole (Fiesole Collection Development Retreat, Fiesole, Italy) recently they were still trotting out the 'fact' that content is king; when in fact the consumer is very much the king now, and we need to know what the king (consumer) is up to. It has to be said that, probably, publishers and librarians know far less about their audience than anybody in the commercial or retail world, and if you want to see good practice in this regard look no further than Tesco. They spend millions of pounds on understanding the end-user.

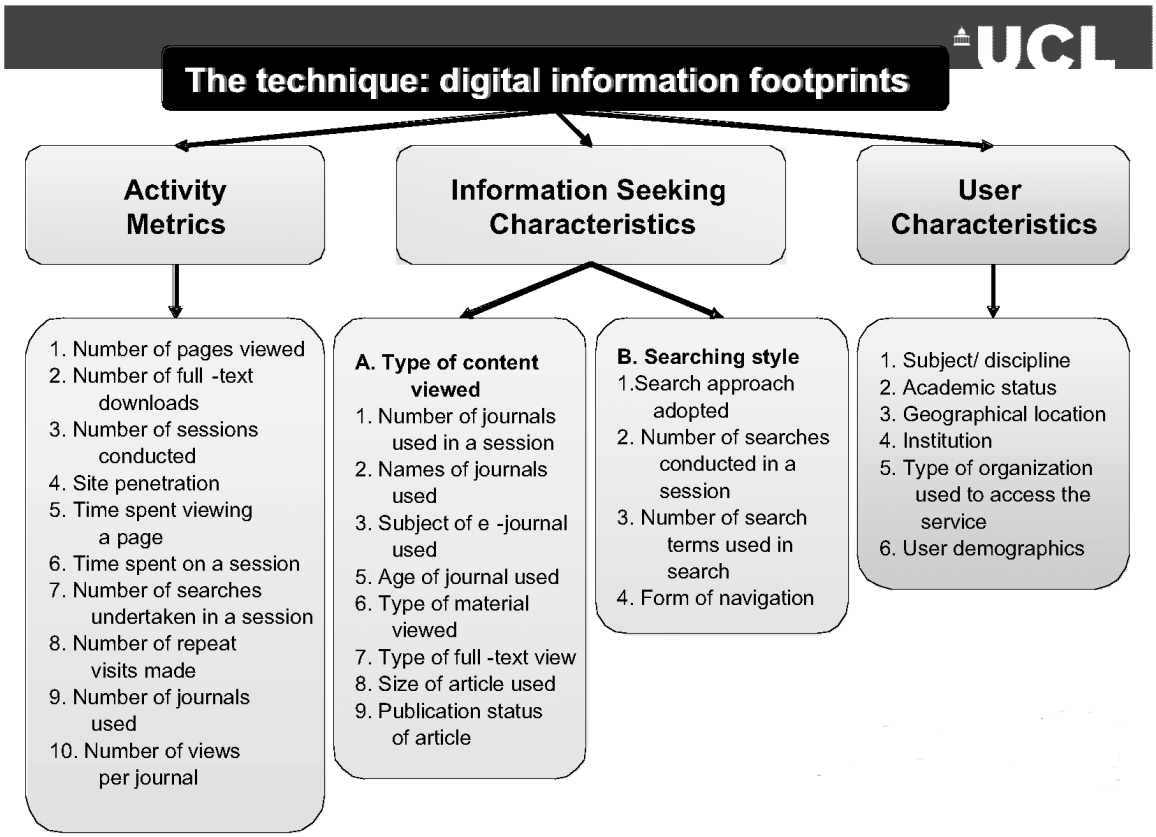
By showing what can be done with somewhat less funds than are available to Tesco, we shall examine what CIBER research has discovered

about information seeking and what can be accomplished by our particular form of analysis, deep log analysis. The research comes from the Virtual Scholar programme that we have been working on for six years. During this period we have undertaken a large number of individual projects and through them developed a comprehensive and robust picture of the virtual scholar. I am going to refer to some of them by way of illustration, in particular the UK National E-books Observatory study which we are currently doing for JISC. At the moment we are sunning ourselves with the 22,000 questionnaires which we have received from the 127 universities taking part in the Observatory. The project is taking LIS (library and information science) research to a new scale and the whole scholarly community is joining in so that we can all understand the impacts of delivering e-textbooks. Three other studies inform this article, one on the impact of open access journal publishing for OUP, another an investigation of the impact and use of journals for the Research Information Network and the third on the behaviour of the 'Google Generation' for the BL and JISC.

In investigating the information-seeking behaviour of the virtual scholar we have drawn from the logs and associated databases a 'footprint' or 'fingerprint' made up from 29 key features, as shown in Figure 1.

What follows is a light-hearted evaluation (but with very serious undertones) of what we have found as a result of applying this footprint to ten million online transactions undertaken by several millions of users.

Most importantly we have discovered that there are huge numbers of scholars and great demand for the scholarly information product. Access is clearly the main driver here. It means that those people already using the system can access this on trains, at airports and in cafés, and this pushes usage ever upwards. It also means than many more people are being drawn into the scholarly net: after all, we are all scholars now. These people are the scholarly equivalent of the expert patient; there are loads and loads of people out there accessing scholarly systems to conduct their own personal research, or someone else's (as is the case with parents). It is a great success story. If we can identify the scholarly outcomes associated with good information-seeking behaviour then we are going to fuel this growth even more.



Key features of digital information-seeking behaviour

There are, then, great, great numbers of scholarly users and in many cases they come from overseas; something which is causing all kinds of interesting dilemmas for government. Thus in the case of a number of UK government-funded scholarly websites we have studied (e.g. Intute, BL Learning), less than a third of visitors were from the UK. Even in the case of the OUP-published journal *Glycobiology*, less than 7% of visitors were from the UK. Yes it is an international journal, and the USA has a very large population of scholars in the field, but 7% for the UK does appear to be low? Our study of the OUP open access initiative is showing that Asia loves open access; information is being transferred at a great rate of knots to create a level playing field. For policy-makers the question has to be, are overseas scholars taking greater advantage of the scholarly resource than UK scholars? Are government-funded investments going to help Americans become better scholars than us? The attraction of the UK scholarly resource is not surprising because the UK has a particularly good brand when it comes to education, therefore scholarly information.

Many users are young. Students represent, in many cases, the majority community for a scholarly service. Some information providers cruelly refer to students as 'noise', as they get in the way (of their prized persistent users), they are very active, but they do not accomplish much. Their information-seeking behaviour is quite different. They spend more time online, viewing.

However, there are even more robots. Robots are the best kept secret of the Internet. At least half of all visits to a site are made by robots. In some cases – Arts and Humanities Research Council (AHRC)-funded 'rarefied' websites – that figure goes up to 96%. What does this mean? That some sites are only ever visited by robots! Robots are quite shrewd and will try and get into a site avoiding the robot dot text box which they are supposed to knock at and state "hey, I'm a robot please send me to places I am allowed to go". They mimic human behaviour, they slow down, they do not consume so much, and Google's robots are the best at this. This way they can go all over the place, to places where they should not go. Should we be happy about this or should we be sad about this, or what?

So what have robots got to do to mimic the human scholarly information-seeking behaviour we have witnessed in the logs? First, they have got to be promiscuous. Around 40% of people visiting a site do not come back, they shop around. We can ascribe this to poor retrieval skills, leaving their memories behind in cyber space, massive choice and Google constantly refreshing that choice. Young people are even more promiscuous.

Second, they have got to 'bounce'. Half of all visitors view one to three pages from the thousands available to them in a site. They bounce in and then bounce out again. They bounce courtesy of search engines and because of choice, shortage of time and the sheer pleasure of bouncing. Overseas scholars bounce less – a picture is building of superior searching by overseas scholars – and young people bounce more.

Third, they should flick. Some bouncing can be attributed to flicking, a kind of channel hopping, checking form of behaviour. I always use my daughter as an example here. She's sitting on the sofa with a remote in her hand watching the TV and she's flicking from channel to channel and I say to her, "Victoria, can't you make up your mind?" And she says, "Dad, I'm watching it all". We are all watching it all, hence bouncing and promiscuity.

Fourth, human information seeking is conditioned by e-mailing, executive summaries and text messaging, therefore robots should not view an article online for longer than two minutes because hardly anybody does that. They should also spend more time reading short articles online than long ones. Quick wins is what people go for. If it is long, either read the abstract or squirrel it away to a day when you will NOT read it.

Fifth, power browse, Hoover through titles, contents pages and abstracts at a huge rate of knots. Abstracts and contents pages are made for that, they are the motorways by which users drive through content. Now with the contents of books opening up thanks to digitalization, it is possible to power browse through them as well. Information seeking has moved from the vertical to the horizontal.

Sixth, spend a lot of time navigating; navigating towards content in very large different spaces is a major human activity. People spend half their time navigating to content. I suspect people like navigating: the actual journey is as much fun as getting there, in fact, getting there is not that much fun because there is reading to do.

Seventh, not everyone is the same. In the huge population of digital consumers there is massive diversity. Students do not behave like staff, women

do not behave like men, chemists do not behave like historians and Germans do not behave like Italians.

Eighth, brand: librarians talk a lot about it, but ignore it as it is almost impossible to determine in cyber space. There are so many players involved, it is difficult to know where authority or responsibility lies. There is though, brand 'cool', which does require thinking about. For instance, many librarians have decided Facebook is cool, a place where they can reach the young. Unfortunately the young are now deserting Facebook, it is no longer cool with librarians camped there.

Ninth, do not behave like a librarian. It was once thought that, with the Internet providing everyone with access to vast stores of information, people would learn to behave like them (i.e. in an organized and evaluatory manner). It has however not turned out like that. In fact, the massive information choice provided by the web and the use of a common platform to undertake a whole variety of tasks, including shopping, has turned everyone into e-shoppers. The virtual scholar and the e-shopper use exactly the same number of words when they undertake a search on ScienceDirect or the John Lewis website.

In conclusion, then, a good deal of scholarly information-seeking behaviour can be portrayed as being active, bouncing, navigating, checking and viewing. It is also promiscuous, diverse and volatile. Twenty years ago nobody would have portrayed information seeking in these terms. But it is true, based on a vast evidence bank, and nobody has disputed it yet. But what does this all mean to information providers and society as a whole? Let us leave it to *The Guardian* to judge. This is what they said in a review of the CIBER Google Generation study, which reported on the information-seeking behaviour described in this paper:

> *The study confirms what many people are beginning to suspect: that the web is having a profound impact on how we conceptualise, seek, evaluate and use information. What Marshall McLuhan called 'the Gutenberg galaxy' – that universe of linear exposition, quiet contemplation, disciplined reading and study - is imploding, and we don't know if what will replace it will be better or worse. But at least you can find the Wikipedia entry for 'Gutenberg galaxy' in 0.34 seconds.*[1]

What is really interesting is that while many people felt that this form of behaviour was associated with young people, few suspected it was also characteristic of adults. We are all the Google generation.

What we have to do now is to determine what these information-seeking traits are leading to in terms of scholarly outcomes. It is necessary to shift our concerns from access to outcomes. Has massive digital access made things better, and how can we help ensure that it does?

References

1. Naughton J, Thanks, Gutenberg – but we're too pressed for time to read. http://www.guardian.co.uk/media/2008/jan/27/internet.pressandpublishing (Accessed 19 May 2008)

■ **Professor David Nicholas**
CIBER
Director, School of Library, Archive and Information Studies and UCL Centre for Publishing
University College London
Henry Morley Building
Gower Street
London WC1E 6BT, UK
E-mail: david.nicholas@ucl.ac.uk
Tel: +44 (0)20 7679 2477