

What are information professionals doing to help users of the scholarly literature?

*Fytton Rowland, Department of Information Science, Loughborough University,
Loughborough, Leics LE11 3TU, UK.*

Abstract

It is important that information and computing professionals give the needs of the users of the scholarly literature high priority as they design new electronic information systems. Early user studies concentrated on human factors applying to each individual electronic journal, and some design principles were established. Recently, as the number of journals available in electronic form has increased, information professionals have begun to provide facilities designed to assist users in retrieving information from the electronic journal system as a whole. This paper reviews some of these facilities and the contribution that they may make to improved usability of electronic journals, but concludes that, despite these advances, actual usage of journal articles in electronic form remains limited.

Introduction

A great deal has been said and written in recent years about electronic publishing and digital libraries (1), and many aspects have been discussed: technical and economic issues; authors', publishers' and librarians' perspectives; copyright and government policy issues. It is important, however, not to lose sight of the key issue. All publications, all information services, and all libraries digital or otherwise, exist to help users. If we do not carry the users with us, our work is in vain. And not all users are like us: they may not be fascinated by the detail of electronic publication, they simply want a usable, convenient tool to help them in their own work.

Publications other than scholarly ones take their readers' needs very much into account. Newspapers and magazines tailor their look and style to the tastes of that segment of the population which is expected to buy the publication. General trade books do not get published at all unless the publisher judges that there will be a sufficient market for that particular type of book. Educational texts -- school and college textbooks -- and their computer-based equivalent, CAL packages, are very carefully designed so that they are pedagogically effective as well as academically correct. Often CAL materials require a team of professionals to produce them: the subject expert, the educationalist, the graphic designer, the computer professional, the video director and the sound editor.

It has often been observed that the scholarly literature differs from other kinds of publication in that the authors drive the system. It is based not on demand from readers, nor on the needs of advertisers, as other types of literature are. Material is published in the scholarly literature because authors wish to publish, and need to publish to assist the development of their careers. Nevertheless, it is usually necessary for authors to be able to demonstrate not only that they have published their research, but also that someone else has read it; usually this is proved by citation analysis. It is therefore necessary to take readers into account as well as authors.

Scholars, however, often do not take into account the users' needs when producing scholarly publications whether as authors, editors or referees. The text may be written in dense, difficult language. The type may be small and the layout unattractive.

Scholarly articles are intended to be an archive of human knowledge, and an archive is often exactly what they look like. It is often anecdotally stated that the average readership of a scholarly article is less than one person. The knowledgeable American commentator Donald King, however, has been collecting evidence about the use of the journal system for many years and argues for a larger readership (2, 3).

We may take the cynical view that no-one reads the research literature anyway, that it is only there for the benefit of its authors, and therefore that it needs to be produced and stored at minimum cost. But if do not agree with this view, we must also take note of readers' needs in designing the scholarly communication system of the future. This paper looks at what has been done to study users of electronic journals, to define their needs, and to help them make use of digital libraries easily and comfortably. Information professionals, and computing professionals too, have a record of devising clever features for information systems that nobody wants and nobody uses. Can we avoid this trap and provide users with features that actually will make their work easier?

User studies

The most obvious way of finding out what users want is to ask them. However, it is always difficult to ask people what they want of some service that does not yet exist. Thus, user studies of electronic journals could not really be undertaken until a reasonable number of electronic journals existed. It is interesting, however, that even the most recent (August 1999) edition of Charles Bailey's bibliography on electronic journals (1) shows a very sparse representation of user studies. The long-established principle of ignoring the user seems to be alive and well in the world of electronic publications.

Nevertheless, my colleague Cliff McKnight reviewed user studies from 1980 onwards at the last ISDL conference in 1997 (4), under the title *Electronic Journals: What do users think of them?* and came to the following conclusions about the use of electronic versions of journals:

- People don't like reading from screens
- People like to annotate
- People don't read journals at their desk
- People don't sit still while they read
- People like to browse
- People don't necessarily want to search
- People like to find things by accident
- People use more than the current issue
- People like stability

Professor McKnight is a leading authority on human factors in electronic publications, and it is barely necessary to update his remarks of two years ago now: they remain valid, so far as use of any one individual electronic journal is concerned. Further user studies have been published in that time (5, 6, 7), but they do not in general seem to suggest any massive increase in the popularity of reading journals articles in electronic form. In the two years since then, however, many new features have come into existence.

In particular, we no longer look at each electronic journal in isolation. People researching in the electronic journals field have been saying for some time that we need a "critical mass" of electronic journals before users will start to use them enthusiastically (8). No-one will want to learn a whole new way of doing things just to use a small proportion of the journals they are interested in. Now, as most of the leading scholarly journal publishers have made their products available in electronic form alongside the paper version, the "critical mass" is being achieved.

Where are we now?

Now we can speak of an electronic journal system, rather than just a certain number of electronic journals. Various projects, especially those within the Digital Libraries program in the USA (9) and the Electronic Libraries (eLib) programme in the UK (10), have looked at the enhancements of the communication system as a whole that may now result from the existence of the critical mass. In particular, of course, there is the potential to link many, eventually perhaps all, journals together via hypertext links within the World Wide Web. Furthermore they can also be linked to the abstracts and indexes (secondary) databases like *Medline* and *Chemical Abstracts*; and many libraries are now linking their full text electronic resources directly to the catalogue systems (OPACs) as well. Other new features such as intelligent agents and push technologies are also intended to assist the user in making use of the entire information system, not just one individual article or individual journal.

At the time when Cliff McKnight spoke here two years ago (4), each publisher had mounted their own stable of journals on their own website, and expected users to log into that site, probably with a password, before proceeding to explore the content of that publisher's journals. It was already becoming clear -- for example, from our Café Jus project (6) -- that this means of access is unsatisfactory. Most readers do not know who publishes their favourite journals, and they certainly do not care who does. They want to be able to scan all the main journals of their own field, regardless of who publishes them, and they do not want to wade through lots of journals in other fields altogether, which happen to be published by the same publisher. Furthermore, they would find it best if all the electronic journals could be accessed from a single interface, preferably a well-designed and ergonomically effective one. The proprietorial attitudes of some publishers are unhelpful here, and seem to show a lack of understanding of how their

own products are actually used by scholars and students. A digital library needs to be like a paper library: it should contain publications from many publishers, arranged by subject classification, not by publishers' names.

Last year when I spoke at the International Joint Workshop on Digital Libraries 1998 in Bangkok (11), some of the tools and approaches needed to overcome these difficulties were beginning to come into existence. Progress is rapid in this field, and the past year has seen a number of further initiatives, and some earlier ones have come to fruition and published their final reports.

Aggregators

A clear trend has been established towards aggregators -- that is, companies who provide access through a common interface to the journals of a large number of publishers. Several of these aggregators are the traditional subscription agents such as Swets, Blackwell's, Dawson's and Ebsco. Others have emerged from the not-for-profit library-cooperation sector, such as OCLC in the USA and ingenta (formerly BIDS) in the UK. Yet others, like Ovid, have developed out of the traditional online bibliographic database hosts. And yet others are publishers themselves, like Elsevier, with its ScienceDirect service, who seek to persuade smaller publishers to use their facilities. Each of these companies is trying to establish itself as a "one-stop shop", just as the subscription agents did in the day of the print journal. Unfortunately, while publishers recognised in the print era that subscription agents performed a useful function and were happy to do business with them, in the electronic era some publishers seemed unwilling to market their products in this indirect way. They expected every library, or even every end-user, to access the publisher's own website directly, as mentioned earlier. Progress has been slow, but it does seem

that in the last year the publishers have mostly recognised the value to the user of a uniform interface, and are beginning to be more willing to allow their products to be marketed through a number of different aggregators.

National or regional licence initiatives

The question of accessing a mass of journals from a single, uniform interface has been addressed in a more radical way in the UK by two initiatives of the Joint Information Services Committee (JISC) of the higher education funding councils. The Pilot Site Licence Initiative (PSLI) ran from January 1996 to December 1998, though the contract with one publisher extends to December 1999 (12); and the National Electronic Site Licence Initiative (NESLI) is now established (13). The PSLI paid four publishers a sum from central funds to ensure that their income from UK higher-education customers remained approximately constant, and this sum was taken ("top-sliced") from the total funding for UK universities as a whole. Then higher-educational institutions in the UK could obtain the journals they had previously subscribed to in printed form from the participating publishers at a discount (in one case, a 100% discount) off the usual price, and also receive electronic access to all of those publishers' journals. The experience of the PSLI enabled JISC to design NESLI, which deals only with the electronic versions and has nothing to do with print-journals subscriptions. Under NESLI, JISC has appointed two managing agents for the scheme -- the Manchester Computing Centre (MCC) for technical aspects and Swets for commercial aspects -- and Swets are seeking to negotiate with publishers. Under the scheme, publishers working through NESLI offer favourable terms for electronic access to their journals to UK higher-education customers; once a deal has been reached with any particular publisher, their journals are presented through a uniform interface at MCC. A

number of publishers have already negotiated deals, and if most do, the desired "one-stop shop" may be achieved for UK academic users. In other countries, voluntary consortia of libraries, regional in the case of larger countries such as the USA, are trying to negotiate similar arrangements with publishers.

Subject gateways

Another valuable approach is the provision of routes into the WWW by subject. Scholars in different fields of research belong to different research cultures (14), and follow different practices in how they conduct their research, disseminate their results, and gather information. In some fields, such as chemistry, the journals literature is pre-eminent; in some humanities fields the monograph still dominates; in engineering fields, conference presentations are more important than journals; and in high-energy physics, the electronic preprint server at Los Alamos National Laboratory (15) has achieved a central position in the communication system.

Recognising this variability, many organisations and individuals have set up web gateways classified by subject. One of the more systematic efforts in this direction has been undertaken by the UK eLib programme (10), which has supported the development of gateways such as SOSIG (social sciences) (16), EEVL (engineering) (17) and OMNI (medicine) (18). These gateways are not explicitly or exclusively concerned with primary publications as such; they provide links to evaluated sources of information of all kinds in their discipline. However the journals of a field are a key source of quality information within that field, and so journals are increasingly linked to the subject gateways. This is particularly true of journals that are free of charge to the user. There are logistical problems in providing seamless links to sources that can be accessed only if

the user's institution subscribes to the electronic journal in question. However, these gateways do, in the case of commercial journals, provide links to the relevant publisher's sites.

Journal clusters created by publishers

One of the major projects within the eLib programme was SuperJournal (7, 19), which was carried out by a consortium of twenty publishers together with several university and research partners. Since a large number of publishers was involved, it was possible to put together clusters of journals within coherent subject areas, and participating university libraries were able to offer these clusters of journals to their users, who found them more valuable than individual journals accessed singly.

Automatic clustering by hypertext links

Many years ago, Eugene Garfield and others at the Institute of Scientific Information (ISI) developed the concept of citation analysis (20), by which the strength of the connection between two different articles could be inferred from the extent of their citation linking. Either the two papers cited many of the same older references ("bibliographic linking") or there were many later papers that cited both of them ("co-citation"). The effectiveness of these measures of connection between papers has been explored extensively, especially by Henry Small (21). The concept is that by giving numeric values to the linkages between pairs of papers, one can group, or cluster, papers together into groups that are presumably about the same subject. This process can be undertaken algorithmically based on the ISI Citation Index databases, and classifies items into groups without requiring any human intellectual classification effort. In effect, the necessary intellectual effort has been undertaken by authors, when they decide what earlier papers to cite in their new paper. However, the process of citation is a slow one. A

piece of research is undertaken, a paper is then written describing it, this paper is submitted for publication, it is refereed and edited, it appears, it is read by readers who are themselves carrying out research, and some of them then cite it in the reference list of their next paper, which itself then goes through the same editorial process. Rarely will a paper be cited in under a year from its own publication date, and more months will then elapse before that citation itself gets into the Citation Index. Thus citation patterns take many years to emerge.

Recently, however (22), it has been suggested that a similar kind of clustering might be undertaken in the era of electronic publishing by analysing the hypertext links between documents on the WWW. The concept is similar; documents that are on related subjects are likely to link to each other, while totally unrelated topics (organic chemistry and Indian politics, say) will rarely link to each other, except due to very exceptional events such as the Bhopal disaster. Webcrawler software can proceed from a site known to be relevant to the user in question (because the user has frequently looked at it) to other sites that are heavily linked to the first. Because the development of information structure on the web is far quicker than print publication, the linking patterns can emerge in months or even weeks rather than years. A WWW search engine named Google operates on this basis, ranking its hits on the basis of the richness of the links to a site. A site that answers the user's query and also has many other sites linking to it, will be preferred to one that also fits the subject description but has attracted links from relatively few others. This approach is not uncontroversial, however, since WWW hypertext links are much more ephemeral than bibliographic citations, and may be thought to be inappropriate as a means of selecting refereed articles for reading by the user.

Intelligent agents

A further step along the road of automating the user's search for relevant papers is the use of the intelligent agent. These software packages analyse automatically the items that a particular user has retrieved in the past, and then try to offer the user other, similar items in the future. This is essentially an automated version of the "selective dissemination of information" (SDI) or "current-awareness service" (CAS) that the early computer-based information retrieval services offered to their customers. Those services required considerable effort on the part of the users, in defining their information needs in terms that the relatively primitive software of the 1960s and 1970s could handle. Skilled intermediaries at the information service provider were also needed. Now the software automatically applies statistical techniques to the user's PC's log to determine what items on the WWW are accessed heavily, and then tries to match these to find other candidate useful items. This approach could in principle be combined with the hyperlink clustering method mentioned in the previous section.

Information-retrieval specialists within both the information and the computing professions have tended to use the model of the purposeful user searching for relevant material. User studies such as Elvyn (8), however, have noted that real-life scholars tend to use a browsing approach instead, looking through recent issues of known journals until something interesting catches their eye. Since this is the way in which printed journals are used, electronic journal systems have to provide a browsing function, and most do. There are misgivings among some users about the intelligent agent approach, since they fear that by becoming dependent upon a software agent, they may miss interesting papers that might have been found by serendipitous browsing. There is also fear that the widespread use of such agents might fragment research communities,

since possibly users will each only see material that supports their own view or position and will not be directed towards contrary views.

Conclusion

In the two years since Cliff McKnight spoke at ISDL '97 (4), the world of electronic journals has moved on from considering each journal, or in any event each publisher's output, in isolation, to deeper thinking about the whole system of scholarly communication.

There is much debate about possible realignments on a completely non-commercial basis (23), but even if we assume that the existing scholarly publishing industry continues, the system will become more interlinked. Users may expect to be able to move around easily between the journals of different publishers, and also between the primary literature, the secondary databases and review publications. There is a potential conflict between this requirement and the wish of each publisher to preserve their own revenue base. It may be that the Digital Object Identifier (DOI), as defined by the American Association of Publishers (24), will assist in keeping track of precisely what each user has retrieved, and who owns it, and ensuring that the appropriate payments are made to the owner.

It has been clear for some time that electronic journals have the potential to incorporate features that cannot be provided in print. These features include interactivity (the ability of users to reply easily to the editor or authors), multimedia features such as video clips, sound and animated diagrams, voluminous research data, mathematical material available directly for further computation by the reader, and rich hypertext linking. Various journals have provided one or more of these new facilities, but so far the majority of journals made available electronically have

been straight copies of the printed version. Another strand of the SuperJournal project in the UK (19) has been to investigate such enhancements. Where the extra facilities have been provided, however, there is not much evidence available to suggest that they are heavily used.

However, user studies that demonstrate that real users actually want to retrieve published, refereed material in a highly interactive way are still lacking. It is clear that throughout the academic world there is a huge amount of use of the Internet for less formal interaction, using facilities like e-mail discussion lists, newsgroups, personal web pages and so on. But when it comes to use of the formal, refereed literature, there still seems to be a lack of enthusiasm for the electronic medium among users. Recent surveys by Tomney and Burton (5) and by the SuperJournal team (7) seem to indicate quite modest usage of electronic journals by end users. Publishers, librarians and computer system developers have all put in a lot of effort over the past three or four years to make electronic journals available to end users. But so far the enthusiasm of end users for them seems limited (25), as indeed the enthusiasm of authors does too (26).

The information professions have been making strenuous efforts to help users to make their way around the new electronic world of journals. Until further user studies are undertaken, however, and demonstrate heavy use of new facilities, there will be some doubt about the effectiveness of this work. Perhaps users obtain the kinds of information that needs heavy interaction and dense linking from other sources, not journals. Perhaps they consciously or unconsciously recognise that the refereed journal literature is more a canonical archive than a dynamic, interactive current information service.

References

- 1 Bailey, C. W. (1999) Scholarly Electronic Publishing Bibliography, Version 26, available at: <http://info.lib.uh.edu/sepb/sepb.html> Updated frequently. Checked 18 August 1999
- 2 Tenopir, C. and King, D.W. (1996) *Electronic Publishing: A Study of Functions and Participants*, in: 17th National Online Meeting Proceedings, New York, May 1996, edited by Martha E. Williams. Medford, NJ: Learned Information, Inc., pp. 375-384
- 3 King, D.W. (1998) Some economic aspects of publishing scholarly journals on the web. *Newsidic*, December 1998, 11-15.
- 4 McKnight, C. (1997) Electronic Journals: What do Users Think of Them? *Proceedings of the International Symposium on Digital Libraries 1997 (ISDL'97)*, Tsukuba, Japan, September 1997, pp. 23-27. Also available at: <http://www.DL.ulis.ac.jp/ISDL97/proceedings/mcknight.html> Checked 18 August 1999
- 5 Tomney, H. and Burton, P.F. (1998) Electronic journals: a study of usage and attitudes among academics. *Journal of Information Science*, **24**(6), 419-429
- 6 Woodward, H., Rowland, F., McKnight, C., Pritchett, C. and Meadows, J. (1998) Cafe Jus: an electronic journals user survey. *Journal of Digital Information*, **1**(3), available at: <http://jodi.ecs.soton.ac.uk/Articles/v01/i03/Woodward/> Checked 18 August 1999
- 7 Dawson, H. (1999) Putting the super into journal: the SuperJournal project at the British Library of Political and Economic Science. *Vine*, **111**, in press. Will be available at: <http://agent.sbu.ac.uk/publications/vine/> Checked 18 August 1999
- 8 Rowland, F., McKnight, C. and Meadows, J. (eds) (1995) *Project ELVYN: An experiment in electronic journal delivery; facts, figures and findings*. London, UK: Bowker-Saur.
- 9 D-Lib (1999) Available at: <http://www.dlib.org/> Checked 18 August 1999
- 10 eLib (1999) Available at: <http://www.ukoln.ac.uk/services/elib/> Checked 18 August 1999

- 11 Rowland, F. (1998) Recent developments in scholarly publishing and their impact on libraries, in *Proceedings of the International Joint Workshop on Digital Libraries*, Bangkok, Thailand, September 1998. *Digital Libraries*, **12**, 47-64. Will be available at: http://www.DL.ulis.ac.jp/Dljournal/No_12/ Checked and found "under construction" 19 August 1999.
- 12 PSLI (1999) Available at <http://www.jisc.ac.uk/progs/index.html#pilot> Checked 18 August 1999
- 13 NESLI (1999) Available at: <http://www.nesli.ac.uk> Checked 18 August 1999.
- 14 Rowland, F. (1999) Diffusion of Information across the Sciences. In *Communicating Science: Professional Contexts*, ed. by Scanlon, E., Hill, R. and Junker, K. London: Routledge, in association with The Open University, pp. 150-156.
- 15 arXiv.org e-Print archive (1999) available at <http://xxx.lanl.gov/> Checked 18 August 1999
- 16 SOSIG (1999) Available at: <http://sosig.ac.uk/> Checked 18 August 1999
- 17 EEVL (1999) Available at: <http://www.eevl.ac.uk/> Checked 18 August 1999
- 18 OMNI (1999) Available at: <http://omni.ac.uk/> Checked 18 August 1999
- 19 SuperJournal (1999) Available at: <http://www.superjournal.ac.uk/sj/> Checked 18 August 1999
- 20 Garfield, E. (1970) Citation indexing for studying science. *Nature*, **227**, 669-671.
- 21 Small, H. (1973) Co-citation in the scientific literature: a new measure of the relationships between two documents. *Journal of the American Society for Information Science*, **24**(4), 265-269.
- 22 Borgman, C.L. (1999) Books, Bytes and Behavior: Rethinking scholarly communication for a global information infrastructure. In *Proceedings of Is there a future for informatics research? A strategy for the 21st century A Cranfield Conference in honour of Jack Meadows. Information Services and Use*, in press.
- 23 Rowland, F. (1999) Electronic Publishing: Non-commercial Alternatives. *Learned Publishing*, **12**(3), 209-216
- 24 DOI (1999) Available at <http://www.doi.org/> Checked 18 August 1999.
- 25 Rowland, F. (1999) Electronic Journals and their Management. *Vine*, 110, 3-5. Will be available at: <http://agent.sbu.ac.uk/publications/vine/> Checked 18 August 1999
- 26 McKnight, C. and Price, S. (1999) A survey of author attitudes and skills in relation to article publishing in paper and electronic journals. *Journal of Documentation*, in press.