Wim Luijendijk

Future Handling of Electronic Journals by Subscription Agencies

Librarians and academic officers began in the early 1990s to explore the capabilities of electronic technologies as solutions to the problem of scholarly communications in print form. The advantages of electronic publications are numerous and have been widely discussed, hence, we will only briefly enumerate them:

- Electronic journals can, in principle, be more current than print journals; their contents can be released to readers as soon as they are ready.
- Electronic journals can be accessed in a variety of innovative ways.
- Arguably, they do not need additional indexing.
- Their contents may be downloaded into readers' workstations for further analysis and manipulation.
- They do not need to be bound or shelved.
- They make minimal demands on space.
- Missing issues and missing pages do not need to be replaced.
- Many individuals can use them simultaneously, eliminating the need for multiple subscriptions.
- They impose no page limitations on authors, permitting extensive supporting information to be included or incorporated into articles via electronic links.
- They can be accessed wherever an institution's telecommunications network or the Internet reaches.
- They permit interactive dialog between, and among, readers and authors.
- They make possible a wealth of value added services, e.g., customized selection of information, electronic navigation within and among documents, the incorporation of multimedia information, etc.

- They present more manageable conservation and preservation problems.

We are in the middle of an information revolution and the latest statistics show more than 108 million people worldwide have access to the Web. There are some 1.9 million homepages that enable shopping, reading of the daily news and the ability to search for an almost endless amount of information.

Publishers

Some major publishers have made their paper journals electronically available and some publishers will allow subscription agencies to function as an intermediary. However, at present some will not, but this may change in the future as on a worldwide basis it will be impossible for publishers to deal with customers directly only already if one looks at the banking function subscription agencies perform for publishers. Publishers will hardly do anything on behalf of their customers without having received a prepayment for their services. Subscription agencies perform this service for publishers and supply credit lines to the marketplace, which in some countries can be quite extensive. I will name a number of publishers which already have most of their journals available via the World Wide Web and we expect initial usage by the marketplace to begin in the year 1998.

Elsevier: Science Direct
Springer Verlag: SpringerLink
Academic Press: APPEAL or IDEAL (Consortium)
Taylor and Francis
Blackwell Scientific
Kluwer Academic
American Institute of Physics
and the number increases weekly.

For instance, the Royal Library in The Hague which is going to be the legal deposit library for electronic journals in The Netherlands, at present has identified 12,000 electronic journals using a storage capacity of 3,500 gigabytes, and they estimate in 5 years time they will have 26,000 journals which will require a storage capacity of 11,000 gigabytes.

A very interesting development I would like to mention here is HighWire Press, located at Stanford University. The university started some two years ago to make sure that non-profit scientific publications did not get left

behind as publishing shifted to the Internet. Now they have some 30 journals online and the most important one is the Journal of Biological Chemistry, which was brought to the Internet in 1995 and is published weekly by the American Society for Biochemistry and Molecular Biology.

The printed version runs about 700 pages yearly and costs some $1,400 for an institution. The online version costs $1,100.

Subscription Agencies

There are still many smaller publishers that are not thinking about electronic journals and at EBSCO we have been working for over five years already to obtain the rights from publishers to put their journals in an electronic file. At present we have over 1,800 journals which are not available directly from the publishers and therefore only available electronically via EBSCO. We therefore differ from other subscription agencies who are operating globally, such as Blackwell's, Dawson and Swets.

In this respect I would like to mention Ovid Technologies, which is also becoming an information aggregator just like EBSCO and they have selected some 300 journals so far in the biomedical sciences. They call their technical format "live text" and Ovid claims this to be superior to page image or PDF format.

Blackwell's has developed an electronic journal system which they call the Electronic Journal Navigator; Dawson has developed a system called IQ, and Swets has a system called SwetsNet, whilst EBSCO's electronic journal system is called EOSS (EBSCO Online Subscription Services).

There is no difference between the four subscription agencies what those systems intend to do, which is basically:

- Integrating the library site's access to electronic journals via the World Wide Web
- Managing electronic licensing agreements with publishers
- Providing support and advice on electronic license definitions
- Dedicated customer support
- Dedicated technical support
- Product and software training materials for librarians and end users
- Combined invoicing for print and electronic journals.
- Subject searching

- Keyword searching
- Browsing
- TOC services
- Alerting services
- Individual document printing

All this may sound interesting and very challenging, but we are facing some real and severe practical problems.

Richard Luce, Director of Carmel Los Alamos National Laboratory Research Library, made an informative presentation to EBSCO in March this year on acquiring, accessing and managing online journals. He went into considerable detail regarding Los Alamos commitment to providing electronic access to their scientific research community. Los Alamos is one of the premiere research institutes in the United States, focusing predominantly on high energy physics. They employ 12,000 people in a 43 square mile facility. They are on the cutting edge with respect to computing hardware and telecommunications. His presentation was on "library without walls" which he characterized as integrated web delivery of information to the desktop. They currently subscribe to 1,800 scientific journals as well as 500 business titles and spend over 2.5 million dollars a year. This represents eighty percent of their total materials budget. He indicated that web based services are a mandate versus proprietary software based systems. In other words, they will not use a vendor who only supports proprietary software solutions.

To give you an idea of the scope of what they are doing, they currently track 5,200 electronic based journals (some of these are self-published items from the Los Alamos physics community network). They have a database that includes 13.5 million citations, 95 million hyperlinks and 200 million cited references. The citations, hyperlinks and cited references are all linked together with their online catalog and this is done through robotically generated lines of software code from a program that they have developed. In developing these links they are guided and driven by the effort to build databases around how people need to get to the information. Everything is integrated with their catalog and citation databases and the bottom line for them is speed. He indicated that their customers (researchers) expect web access with links and integrated with the library catalog so that they have one place to go and one common user interface. They expect everything to be hyperlinked, with new associations and easy to view on the screen with

printing capability. Their user community also wants an alerting service based upon their own personal interest profile and they expect to be able to search and retrieve full text from their databases regardless of who the publisher is. A checklist of concerns he reviewed are as follows:

- Inconsistent methods of access;
- Timeliness - receiving electronic versions at the same time or ahead of print versions;
- Unreliable service - poor response time;
- Passwords - he flatly stated they won't work;
- Usage data - indicated this is difficult to get from publishers so they are developing their own supplier software program in order to get the information they need;
- License agreement - this is a problem and needs to be simplified;
- Article level identifier - this needs to be standardized and his preference is SICI (the Serial Item and Contribution Identifier); however, some publishers, among which are the American Chemical Society, the American Institute of Physics, and Elsevier Science, are heavily marketing PII, the Publisher Item Identifier, whilst lately the American Association of Publishers (AAP) issued a request for proposal to develop a scheme for identification of electronic articles and they call it the Digital Object Identifier (DOI).
- PDF files not fully processed - some PDF but some TIF images with PDF wrappers so unable to take full advantage of PDF searching;
- Rich header information - he thinks it is ridiculous to have to pay OCLC to get this information and that publishers should provide it as value added service;
- Simultaneous user agreements don't work on the web - they are outdated as there is no way to determine simultaneous usage - his view is anyone offering simultaneous user agreements for web access need to rethink the model. In this regard the biggest problem he has is negotiating with suppliers who don't understand the scope of their operation and have

models that won't work for them (they have 17,000 machines at Los Alamos that need access to data).

Another series of concern was uttered in February 1997 in the publication Against the Grain by Clifford Lynch of the University of California.

- Authentication - The problem is with site licenses which guarantee rights on behalf of an institution's users for access to materials on some Web site. How do the Web site administrators know that the users trying to access their material are in fact affiliated with an institution that has licensed rights to some or all of the Web site content? The answers right now are basically unacceptable. They are user passwords or restriction by IP source addresses, basically whatever physical network to which these people happen to be connected. Given that many people do their work from home, given that many people are now using commercial Internet services to get access, and given that we are moving into mobile environments, IP address filtering simply isn't going to work.

- Printing - We need to think through printing, given that most people basically are going to be using this digital information to decide what to print. Printing items by bringing them down to the user's machine and out again as you would do with a Web browser is not necessarily the only process you are going to want. This is particularly true if you are dealing with some form of bit-mapped representation that is going to be painful over a slow line, and which will tie up large amounts of disk space in transit. You are going to want routing both from the workstation and directly from the publisher server to third party printers on the network.

- Content consistency - There is a real problem with how we think about journals in digital form and Web sites. Here is a concrete scenario: You put up a Web site; you get some users in; you show them the journal; you say: "How do you like it now that it is electronic?" The users say, "Oh yeah, this is pretty good." But the content provider has only changed the format of a journal, and that is what the users are reacting to. The publisher hasn't faced the users with the idea that the whole system which they use to access the literature is changing in fundamental ways. This system of access defines how users locate content, navigate within that content, and use it. Different publishers will organize and

present content differently. Different publishers may offer different ways of access to content.

- Linking technology - Clearly we need better linking technology. One promising tool is the new serial item and contribution identifier standard that is about to be published, which essentially lets you produce keys computationally from an abstracting and indexing record or from content and then potentially match on those keys. You could potentially use a real-time, interactive, computational process rather than hard (precomputed) links that way.

- Individual-object-addressing - Many of the sites that are being set up are not individual-object-addressable. This is a crucial point which seems to have eluded most publishers. As a user, you want to be able to point into an object (like an article) on one of these sites so that when you follow the link you go directly to a specific article - not so that you follow it first to a welcome page that gives you the news of the day, offers you discounts on some new journal, offers a trial issue, and only then presents you with a long list of journals from which to select and a choice of which year and issue. You need to be able to make links directly across sites down to the level of individual pieces of content. Clifford Lynch believes that that is an absolutely critical design criterion as we start moving content to the Web in this environment. And obviously, until we get individual-object-addressability (or if we don't get it), making meaningful links in abstracting and indexing databases is going to be a problem. This is also an essential prerequisite for linking from a citation in one article to the text of the cited article, which is something that users really want.

Last but not least, the problems of access as well as archiving if one looks at the tremendous volume of bits and bytes which will be accumulated on an annual basis, will not be easy to deal with.

If the information is stored on a floppy disk, optical disk or tape, one must first have the hardware and software to accommodate the medium such as an optical reader or system with appropriate disk-to-tape drives, and the correct version of the software program used to store the data. If the information is stored in a remote location, but is accessible through software, one must have the necessary software to locate and retrieve or download it. Additionally, one must have the necessary software to access or read the data. This introduces many problems and questions. Electronic journals, just

like print journals, are produced with a variety of software programs that run on several different hardware platforms. As electronic journals become more sophisticated to include photos or graphics or even become multimedia, the number of formats will likely increase. Many programs or formats for these journals will come and go, just like word processing products have done and continue to do.

"Technology refreshing", whereby electronic information is transferred from one waning medium to an emerging one, will need to be practiced. It will be a challenging task for the group or person who wants to archive and provide easy, universal access over the long run.

The Information Age has done much to change the way we look at our industry and the library profession and the way in which we deliver information. Changing times call for a changing mind set.

But all the new technology which we are developing must be with the end user in mind as otherwise it will become useless. The problem we are facing at present is whether we can construct a system that users will find comfortable in making broad use of electronic journals as a substitute for printed ones, not only for today's information, but for all the information we have produced and will be producing from now onwards until many generations in the future.

***

References

1. Malinconico, Michael. Electronic Documents and Research Libraries. IFLA Journal 22 (1996) 3.

2. Luijendijk, Wim. Archiving Electronic Journals: The Serial Information Provider's Perspective. IFLA Journal 22 (1996) 3.

3. Luce, Richard. Acquiring, Accessing and Managing Online Journals. Presentation at EBSCO's Academic Seminar, The Netherlands, March 1997.

4. Lynch, Clifford. Technology and its Implications for Serials Acquisition. Against the Grain, February 1997.