

Non-standard users: The Library

Raf Dekeyser, K.U.Leuven (Belgium)

First of all, I would like to point out that I would not classify the research library among the "non-standard users" of the Internet. The use of the library as the most important source of information about previous scientific research has since long turned out to be an essential instrument for the academic community, and the way in which they use the network has to be seen as an essential ingredient of the general usage of this medium by the research community.

The library has to be seen as a basic service centre for the benefit of better research and teaching. Some people have predicted that libraries will disappear, becoming superfluous with the electronic revolution. Personally, I believe that - on the contrary - due to the increasing availability of information sources, libraries will become more important, but their role will change. They will be less visible (as a physical institute), but they will be developed into a strong back office service for information provision.

Libraries are using the Internet network for several different purposes, which I would like to classify in the following categories: cataloguing, inter-library loan (ILL), database access, electronic reference service, and distributed archiving.

Cataloguing

The development of computers led naturally to the electronic version of the old card catalogues. This evolved into large integrated library management systems, including - apart from cataloguing - also modules for circulation (user control, lending...), acquisition (ordering, budget control...), periodicals (detailed holdings description, sending reminders for missing issues...). These systems evolved first from mainframe applications with dedicated terminals into client/server software, and lately into thin-client applications or direct web-based access systems. Network usage for these applications is, e.g.:

- The installation of one central system for a number of different libraries. There are many important examples of university library networks in this sense, but especially in the sector of public libraries there exist networks of several hundreds of libraries, running on a single central system. Some have mobile (bus-based) units, linked in a wireless way to the system.
- Copy-cataloguing: Since cataloguing is very labour-intensive and many libraries are buying the same book, it is worthwhile to exchange catalogue records. Records can also be downloaded for a fee, e.g. from the Library of Congress in Washington.
- Searching the catalogue: Users (and often also outsiders) can browse or search the catalogue of most university libraries. Registered users may have the possibility to make a loan reservation (or renewal) through the web.
- Collective catalogue building: There is quite some interest in large collective catalogues, e.g. national catalogues, or international ones (such as WorldCat from OCLC). The traditional way of producing such products is by merging the catalogue files of the participating libraries and deduplicating the records (while keeping all information concerning the holding libraries). The result is then a new database, which has the disadvantage of being quickly outdated, so that the whole procedure has to be started all over. A more efficient approach is to build only a virtual joint catalogue, in which each search operation is translated into a simultaneous search through the participating systems. Alternatively, a central index can be built and regularly updated through the harvesting of metadata from the participating catalogues; the harvesting protocol OAMHP from OAI (see below) may be useful for this purpose.

For most of the points mentioned, it is very important that catalogue systems follow strictly internationally agreed standards, both for metadata description (MARC21, Dublin Core...) and exchange of information (Z39.50, e.g.).

Interlibrary Loan (ILL)

ILL is profiting from the Network in the following ways:

- Searching each other's catalogue and requesting a document. The result may be a loan (in general of a book) or the mailing of a copy (often of a chapter or a journal article).
- Accounting: The bookkeeping of the fees that have to be paid between libraries may be calculated automatically (e.g., on the basis of the number of pages for which a copy is requested).
- Electronic delivery of a document, either to the borrowing library or directly to the end-user. This is most easily done when the lending library has already access to the electronic format of the requested document, although most publisher's licenses do not allow to do this. This means that almost always one will have to scan the paper format.

In view of the ease of the electronic exchange, there is an increasing shift from local lending networks into large international ones (e.g., through the British Library's Lending Library section).

Database Access

For most present-day students and researchers, database access is clearly the most important web-delivered library service. The data can be either the already mentioned catalogue records, bibliographic references, full-text articles and even books or course material.

- Bibliographic databases are most often the electronic version of previous abstract or index journals that used to be published in paper format. Searching in the digital version of a database is of course much more efficient than in the paper journal. In the initial stage, many of these electronic databases were searched through dial-in to special service providers, the smaller ones came available on CD-Rom for stand-alone or network installation (with access sometimes shared among libraries), but the latest trend is certainly towards direct web access to the publisher's server (with the obvious advantage of ease of updating).
- Full-text e-journals: This is the area where we have seen the fastest growing revolution during the past years. About 5 years ago, many librarians still did not believe in the possibility of success for this medium. At present, our users would not want to do without their e-journals, and many libraries would be happy to cancel all paper subscriptions in order to take advantage of the price reductions for e-only; it is only the unfavourable difference in VAT-rates that holds them back. The general trend that we observe here is one of centralization. Through mergers between publishing houses, we see electronic access to large title collections being offered to libraries (and increasingly to library consortia). Elsevier is now with ScienceDirect the owner of more than 1500 titles, and aggregators like EBSCO offer databases with titles from different smaller publishers. Through big retrodigitization projects, they are increasing their digital archive collections at full speed (most often in PDF for visualisation purposes, with XML as an archival format for future convertibility). In order to placate the worries of librarians about perpetual guaranteed access, back-up archives are being established, e.g. the Elsevier archive at the Royal Library in The Hague. Special services are set up by these commercial services, amongst which we can mention Scirus, a free search engine through the whole of ScienceDirect but simultaneously also through a huge collection of academic websites, and CrossRef,

a subscription based database that links each paper in a commercial e-journal to its respective website.

Below, we will describe another model, which is much less centralized but based on distributed repositories.

- Books, theses and courseware: With e-books we are apparently in the same situation of wait-and-see attitudes as was then case 5 years ago with the e-journals. Nevertheless, several initiatives are coming up from universities who systematically make their dissertations available on the web, and the initiative by MIT to make its courseware freely available has been greeted with much interest. The digitisation of rare books and cultural heritage documents offers a unique opportunity to make these items broadly available, where their fragility and preciousness before prohibited this. Small-scale university presses, on the other hand, can use the digital medium to achieve a broader commercial distribution of their publication, e.g. through collaborating in a network of "printing on demand", where sold books are printed at the buyer's site in order to save on handling and mailing charges.
- Under the heading of databases, I would also like to include several new possibilities for factual databases. There exist many examples of scientific databases (e.g. detailed outputs of experiments or calculations, genetic codes, results of social surveys...) that deserve being preserved for possible future reuse. In some instances, the directly involved scientific community is taking care of its archiving, but there is a danger that they may loose interest when their first use of the data is finished. Libraries and librarians traditionally attach more importance to the preservation of such materials, and it may be worthwhile considering to charge the libraries with the task of preserving these digital collections, just as they preserved in the past their paper collections. The ultimate aim is to build together a large digital library of scientific knowledge.

Electronic Reference Services

Many libraries have always maintained a reference service for their customers, where questions of any kind could be asked, and where one could at least expect to be guided towards a place or a reference for the correct answer. In the digital age, this has been transformed into services like "Ask a librarian", with many advantages like distant access for the customer, and a face-saving delay possibility for the librarian, who often has to look for the answer himself. In view of the success of such services, the Library of Congress and OCLC started with "Question Point", a network service for knowledge sharing between libraries and other institutions. Questions asked by a customer may be automatically routed to the site most likely to offer an answer, the selection being made on a collection of subject-based strength profiles of the participating libraries. Similar networks are already operational among a large set of public libraries in the U.S., with services offered 7 days a week, 24 hours a day.

Alternative models for scholarly communication

During past years several new models for scientific exchange of knowledge have been proposed and many experiments have been started. Discontent with the present commercial model, with its ever-increasing prices for the scientific journals, is certainly at the basis, together with the rising awareness that in the digital age universities and research institutions can deliver themselves most of the services of the publishing process. Examples like the physicists' Los Alamos "arXiv" of Ginsparg and the PubMed Central database of the National Institute of Health may have acted as a source of inspiration. Actions like the "Public Library of Science" and the "Budapest Open Access Initiative" have further stimulated people into take these initiatives seriously. Whereas the arguments pro and contra these alternative models are not of much relevance for the discussions in the present workshop, we should nevertheless be aware of what is going on, since it may have a strong

influence on the way in which the exchange of scientific communication over the network will take place in the years to come.

The essential ingredient of the model would consist of a network of digital repositories (also called "e-print archives"), either subject-based or institutional (e.g., one per university). Authors should post their publications in these freely accessible repositories. In order to allow an easy access to the metadata of the deposited papers, the Open Archives Initiative (OAI) group developed a standard protocol for the exchange of such metadata in a simple format, the so-called Open Archives Metadata Harvesting Protocol (OAMPH). One of the metadata is, of course, the web address where the full text can be found. In this way, it will be possible to set up "harvesters" that collect these metadata. The University of Southampton has developed free software for the implementation of OAMPH-compliant repositories (eprints.org), and a couple of harvesters are already operational. For the user, they consist essentially of a search machine which points to the web sites of the documents found to correspond to a given search profile.

The main advantage of the described model is that it would offer a financially competitive alternative to the present commercial channels of scientific communication. The main cost of the dissemination is carried by the author's institute, which runs and maintains the repository. The main problem left is that of the validation method that should certify that the information is scientifically correct. Whereas the first archives were just digital preprints, with the standard peer review performed by the commercial journals, the new system allows eventually more innovative methods of validation and impact measurement. Some librarians dream of involving the learned societies or existing editorial boards in this process of peer review, e.g. through new virtual overlay journals (a set of quality labels given by the reviewers to a collection of documents in the repositories). Submitting a publication to such a review process could be subject to the payment of a small processing fee. For the moment, a foolproof quality label management system for such peer review is still missing in the OAI protocol.

Requirements from the library for the Network

Whereas traditional library use is at present reasonably well served by the existing network facilities, we see, from the above discussion, several possibilities where improved bandwidth might become necessary:

- o We will probably see a growing need for simultaneous search and retrieval access to very diverse database systems. Protocols like OAI and Z39.50 already allow access to a distributed collection of databases, with different formats and standards, as long as some conventions are observed.
- o Especially important for Europe (and more so for the Humanities than for the exact and biomedical Sciences) is the diversity of languages. We hope that the future will bring us faster, more powerful and more accurate automatic translation software, which would give a completely new life for all kind of non-English digital books and journals. This would certainly increase the demand of full-text access to digital databases from the Humanities, which for the moment is still lagging behind the Sciences.
- o Most bibliographic databases for the moment allow the search through an extensive set of metadata, including often the abstract of the publications. There will, however, be an increasing demand for full-text searches, inspired by the success of the Google-approach. Searching through the vast amount of scientific publications, present and past (through retro-digitisation), be it on the servers of the commercial publishers or on the servers of a network of e-print repositories, will certainly lead to a huge increase of data traffic on the network.
- o As an example of what may be expected, we refer to the recent search through thousands of repositories performed in the UK, where half a million student's

essays have been checked against 800 million websites and archives for possible plagiarism. (reference: The Times of 8/1/2003 and http://www.ananova.com/news/story/sm_737513.html).

- o It should be clear that the library's collections will in the digital age more and more evolve towards multimedia repositories, where not only text material will be stored, but all kind of multimedia files, such as sound, images, video, etc... Storage and transfer of such data will also necessitate more powerful hardware and more bandwidth.