University of
Strathclyde
Glasgow

Strathprints Institutional Repository

Law, Derek and McSean, Tony (1998) *Net-knitting: the library paradigm and the new environment.* In: 6th European Conference of Medical and Health Libraries, 1998-06-22 - 1998-06-27, Utrecht.

http://strathprints.strath.ac.uk/

Net-knitting:the library paradigm and the new environment

Derek Law   **&**   **Tony McSean**
King's College London   **British Medical Association**
derek.law@kcl.ac.uk   **tone@bma.org.uk**

It is the purpose of this paper to argue that librarians have been blinded to its basic flaws by the gaudiness of the Internet and that we are confusing sources and resources. The Internet shows none of the features required for scholarly communication and whether or not we believe this will change, we should be developing models which offer electronic services as a viable and reliable resource.

Although the Internet is of some age in the dog years which pass for computing time, the World Wide Web is relatively new, with the first web browser dating only from 1994. In the four years after that it achieved a phenomenal acceptance, in what Paul Evan Peters called the largest mass migration in human history. It was adopted by fifty million users in fifty months. Radio took 38 years to gain such an audience and television some thirteen years. Currently it has some seventy million users. And yet it lacks the important elements of sustainability necessary for scholarship:

- Permanence
- Availability
- Accessibility

The Web is in fact a four year old experiment, not a robust service. Not for nothing is it called the World Wide Wait. Not for nothing has a Dilbert cartoon appeared noting that all the time saved through automation and computers in the last fifty years has been entirely outweighed by people sitting in front of pc's waiting for web pages to load. A variety of issues reflect the very real difficulties of the Web for scholarship.

Identifiers and Naming. The continuity of citation is central scholarship. In a print on paper world we take it for granted that a scholarly paper can cite Vesalius, Lister or Suzanne Bakker and that other scholars or libraries can trace and find these publications or data. That stability does not exist on the Internet where there is a basic need to reference objects as they move and change over time and place. Rather resignedly we simply not the impermanence of URLs. Nor is their any consistency over who may name objects since it appears that anyone can. This in turn removes one of our marks of quality. Authorship, ownership and impartiality are easily disguised. Worse, there is no expectation of who will maintain naming over time. It is our experience that libraries, even national libraries are as guilty as anyone else in having created a fluidity that sits uneasily with scholarship. Commercial publishers have created Digital Object Identifiers, but it is not at all clear that they are usable in the very substantial area of primary sources and grey literature.

Metadata. This appears to be under better control by our profession as the Dublin Core now involves Europe, USA and the Pacific Rim. But there is still basic work to be done on how to describe new genres of multimedia and how to describe new services. What kind of record are we to create for a changing and dynamic resource. We also need to describe the terms and conditions of use and at present these vary between locations, between user groups and over time. For example, an electronic textbook may be licensed to be available only to first year anatomy students in the summer semester.

Authentication. This is important for electronic commerce, but there are no good ways of proving membership of the data club when away from home. This is perhaps no more than irritating. More importantly once membership is established through some form of individual log-on, are we willing to give up the anonymity of the user? Most libraries consider it a matter of professional ethics not to reveal who has used what library material, unless a criminal offence has been committed. Some users, particularly those working with pharmaceutical companies, regard their library use as commercially sensitive. Are we really willing to cede this anonymity to gain access to electronic data?

Distributed Search and Indexing. This remains a very big issue with a great deal still to be done technically. Web indexing systems are breaking down as their architecture collapses under the weight of data. A simple quite specific search will frequently produce over a million hits, listed in no discernible order. There were 320 million web pages on the Internet in May 1998, of which no more than 34% are searched by the best search engine – and there is no readily available way of discovering which 34%!

Rights Management Systems. These are being designed largely at the behest of commercial organisations in ways which mirror the power and needs of the entertainment industry. And yet from the scholarly perspective there are at least three major areas of philosophic contention.

– Privacy. As mentioned above, the right to anonymity is both an academic requirement and has been an obligation from the library to its users. Not only is such privacy under threat, there is the further possibility that usage information could be sold on to third parties as marketing information.
– Preservation. Publishers have never had a responsibility to preserve their publications, yet we have no general legal deposit for electronic publications, even if we had a definition of what constitutes an electronic publication. As publishers typically lease rather than sell electronic data, such material must be considered at risk. In any case the technology for preserving electronic material is far from robust. Who is to preserve what remains a major, undecided issue.
– Fair Use. The concept of fair use for private study and research is an important one. Yet rights management systems which prevent general browsing take away that right. Commercial publishers feel it inappropriate to an electronic environment; scholars might beg to differ.

Network Topology. In Europe, at least since the time of the first Bangemann Report this has been assumed to be a matter for the commercial marketplace. Yet scholarship, unlike commercial markets, is both global and goes into uneconomic areas. Problems arise at

both extremes of need. On the one hand high technology scholarship demands very high bandwidth computing at the leading edge of technology. Yet the report on the (in academic terms) quite modest Ten-34 Project to link European Research Networks found that such links were not available commercially. At the other end of the spectrum, the commercial marketplace will not put adequate technology into non-commercial markets. That is to say that there are large parts of the world where networks and network services will lag impossibly if left to purely commercial motives.

Preservation and Archiving of Electronic Information. This is commonly acknowledged to be one of the hardest of areas to resolve for all stakeholders, one fraught with technical, legal and operational problems. Some technical preservation centres have been running for over twenty-five years. They have produced no magic solutions and little comfort in proving that the technical problems are very difficult and very expensive to resolve. Some progress is being made on electronic legal deposit where useful dialogue has opened up with the publishers. But it must constantly be re-emphasised that much of the material we will wish to preserve is non-commercial. Nor is it self-evident who should conduct the preservation. The national libraries might manage the process but it seems safe to assume that issues of institutional continuity will be even more important than in the paper environment, where company take-overs and bankruptcies, incompetence, indifference and even malice have put many historic collections at risk over the years.

Instructional Media, Courseware and new media. Libraries need to rethink their roles and mission in relation to electronic material. This applies both to external material and to internally produced material. In the case of external material librarians must consider not just purchasing licences to make titles available (or ignoring them while departments make the purchases). We would argue that they must consider the relative costs and merits of remote access, local mirroring, consortial purchasing and so on. And this must sensibly involve the total cost including network charges. Even for "free" sites such as pre-print archives it is important to establish whether the archive has a more appropriate European mirror site which turns it from a variably available source into a reliable resource, saving the organisation many hours of waiting for screens to load. In terms of access how far should library staff surf the web to catalogue and record useful sites or e-journals and start to provide information on what is available rather than what is possessed.

Internally the Library has to clarify its role in relation to electronic information. Is it the institutional provider and/or archive for all instructional material created locally? If so, how far does the remit extend? Does it include, for example, collaborative data analysis and its records or knowledge representation and its re-use? Even if the library is not meant to cover this type of activity does it or should it have a role in ensuring that standards are met by the organisation as a whole for issues ranging from standards used to intellectual property rights.

The new media require a major redefinition of the library's role. Even if the outcome is to leave the libraries role as it was, at least the organisation will have ensured

that it has a series of policies and responsibilities in place for dealing with electronic material.

Scholarly Communication. There is a real threat of what has been called cybercolonialism, the overt or covert preference for one set of resources over another. This is compounded by our own willingness to confuse sources and resources. As an example of this it is worth examining a language neutral discipline such as mathematics and comparing the treatment of major and very longstanding east European journals from universities such as Cracow or Warsaw. This will appear in European gateway sites but not North American ones. This problem is worsened by an increasingly common practice – often fostered by search engines – of preferring to use American websites. These may provide better or richer sources but if slow to the point of unavailability for much of the European day are in truth inadequate resources or services.

Two key issues are standards for version control and mirror sites. Originators of data, jealous of its quality are often unwilling to entrust it to third parties without prolonged negotiation. And yet mirror sites are a very economic method of improving network performance. An obvious solution is a set of standards or kite-marking to indicate the quality of potential mirror hosts. This would reassure not only the data supplier but also the data – user, who must at present also rest uncertain of the version of the document which is available.

It is also regrettable that the debate on electronic publishing has been so dominated by the STM commercial model. STM publishing while undeniably important represents only a fraction of the annual acquisitions of most universities. Even in scientific libraries significant quantities of non-commercial material or small learned society material are acquired and this may be expected to grow in an electronic arena. And yet it is not self-evident that systems and practices designed for electronic commerce sit comfortably with the needs of scholarly discourse. Yet at present very little thought is being given as to how we support the scholarly infrastructure of the small learned society or the science and medicine of developing countries.

There is a further category of material at risk, what Clifford Lynch of CNI has called endangered content. Computer Science, a discipline founded and maintained on a non-printed tradition has reached a point where its pioneers are retiring and dying. As a discipline it has only just begun to realise how much of its common heritage it may need and may already have lost. Great efforts are now being made to salvage this position. It illustrates perfectly that in an electronic environment new thought must be given to how we record and locate a discipline. We cannot wait for paper archives to arrive on the death of great men, nor are laboratory books now the only source of laboratory data. A complete reappraisal is needed of how primary research data and even the e-mail of scientists, of bulletin boards and discussion groups is to be maintained to show the traceable path on which science depends.

Network Topology. As already remarked, the United States becomes a virtual country for most of Europe in the afternoon, as the bandwidth clogs and slows with traffic. It is claimed that costs are dropping and so we can simply buy more bandwidth. It is more

likely that the UK experience is typical. For the UK academic community, the cost per bit of international traffic halves each year - but the traffic trebles, inexorably increasing the bill to a point where restraint has to be applied. It has been interesting to see the reaction to this in Australia where costs are passed on directly to universities. In 1998/9 Australian Vice-Chancellors have introduced scheme in which hits on non-Australian web-sites will be charged at twice the rate of local web-sites. This clearly recognises the distinction between sources and resources and aims to manage traffic in sensible ways. We may expect others to follow this model.

Although Australian universities have adopted a model for managing network topology it is not clear whether that model has any theoretical underpinning. We would then wish to suggest that the current model of unbridled access to the anarchy of the Internet is not the only or the best model for managing electronic resources. Intranets and/or regional networks are being created which form more appropriate boundaries for electronic resources. As the cost of filestore drops very quickly it can often be shown to be economic – not least of time – to mirror resources on the local network. Further, the ease of access to electronic resources allows us to revisit the issue of access versus holdings strategies – where access has held sway for some time as a professional dogma – and argue that holdings strategies may again be appropriate in an electronic environment. When managing information, what organisations have never done is line all employees or students up (at least metaphorically) at the start of the financial year, give them one thousand guilders and told them to acquire anything they liked that might help with their work. Instead they have identified the material relevant to the work of the organisation, collected as much of it as they could in one place, employed professional information specialists to manage it and make it available, and arranged controlled access to the information which cannot be held locally for those who can show they require it. This does not prevent individuals using other channels from public libraries to bookshops for any other information they choose. We would then argue that this model of the library provides a perfect paradigm for the management of networked resources.

LIBRARY CORE ACTIVITY	INTRANET CORE ACTIVITY
Acquisition	Resource Discovery
Collection Building	Local Fileservers and Mirrors
Classification	Knowledge Management
Preservation	Long-term Data Sinks
Inter-Lending and Document Supply	Controlled Internet Access
User Instruction	User Instruction

We would contend that there has been an all too ready acceptance of the Internet. Its undoubtedly huge impact on the availability of current information has dazzled us to its flaws as a medium for scholarly communication. It is important to revisit the information needs of our organisations and of scholarship and to re-interpret them in the light of the possibilities and limitations of networks. The model proposed above then describes how in a local context we can take control of the environment and use it positively, acting collectively to meet institutional need.