ELECTRONIC INFORMATION: ACCESS, CONTROL AND AVAILABILITY

Derek Law

Paper given at the 16th UKSG Annual Conference, Southampton, 22-25 March 1993

The title of this conference paper, *Electronic information: access, control and availability* is very plain and straightforward, albeit fairly prosaic, and these three themes of access, control and availability will be discussed in that order. It should be clear that the paper is concerned with nationally and internationally networked electronic information and not the various forms available at local or campus level, another whole topic in its own right.

Access

A quick overview of what is happening in the network area is a necessary preliminary. One should look first at the UK and inevitably at JANET and its successor SuperJANET. In ten years these have moved from being the specialised tools of scientific researchers, to the everyday tools of staff and students in every discipline and indeed well beyond the universities. (1) There is a widely held misconception that JANET is the preserve of the academic community, second only to that other misconception that JANET is subsidised. Both notions are quite false and, indeed, the shifting structure of JANET and its new limited company UKERNA (the UK Educational Research Networking Association) means that it will be even more open to the public. The intention is to try and make the network available not only to other publicly funded bodies such as schools and hospitals and local government, but also to researchers in industry and commerce as well as commercial services. Indeed some such as OCLC and Blackwells are already on JANET. JANET, like all networks, has gateways to other networks ranging from BT's Telecom Gold, to the new EUROPANET in Europe and on through to the Internet. As readers

Derek Law is the Librarian at King's College London, Strand, London, WC2R 2LS

of computer magazines will know, the American-based Internet is now available to the meanest of modems - and also to the greatest; some of you may have noticed that both President Clinton and Vice-President Gore have opened public mailboxes on the Internet. Not very surprisingly anguished cries have begun to emerge asking how to manage a heavy daily influx of e-mail. So everyone with a terminal and a modem or access to a public network is already in business. An article in the April issue of *Personal Computer World* explained how anyone could join and use the services. (2)

The real problem is not then gaining access to the networks but finding a way around them. The next major task that we face as a profession lies in creating the navigation and filtering tools, the knowbots, which will promote access. It is something of an irony that those at the computing frontiers are discovering that their major problem is not technical, but good old-fashioned cataloguing and classification, describing what is available and where it is. As Lorcan Dempsey has described it, "what we have is a flea-market and what we need is a department store."(3)

Some of you may be familiar with names like Alex and Archie, Veronica and World Wide Web, with WAIS and Gopher. It is the Internet Gopher which is exciting most interest at the moment. It is one of the first client-server network information retrieval tools. The world of information is presented in a series of hierarchical menus, but that information could be on one or many computers. This will be transparent to the user. It is even possible to place bookmarks and return to a piece of useful information the next time. These tools will no doubt come to seem very crude first approximations, but there is a great deal of activity in this area and it is very promising. If there is a real concern here it is that librarians are not sufficiently involved in the design process and are leaving too much of the work to computing specialists.

In terms of access one of the largest areas of difficulty we face is that of copyright. Thus far I detect no real signs of common ground amongst any of the parties involved. Nor do I have any easy solutions, so perhaps all I can do is caricature some of the last ditches into which the protagonists seem inadvertently to have fallen.

First come the producers, our dearly beloved academics. Some of them make money from the process, others seek fame and glory - or at least a grade five rating. Most of them have no idea of the intricacies of copyright law and in most cases do not even know that they are signing away their rights on publication. Most of them show a healthy disregard for copyright, at least when faced by a photocopier.

Second come the publishers. To a degree they are terrified by the new electronic world. They are faced with the need to make expensive choices - and the possibility of expensive mistakes. Their role in this new world is unclear and they need to redefine the added value they bring to products, since this is not always obvious to the rest of us. At the same time the Publishers Association pontificates unhelpfully that blanket licensing of networked products is so awful that it is simply not a matter for discussion. This mark you at a time when the normal form of agreement in higher education is the blanket licence.

Then there come the higher education policy makers, breathing fire and brimstone and talking of catastrophe theory. Retain copyright within the universities as a contractual obligation is one cry; start our own journals and cut out the middleman is another. There is a perception that we pay for the research, give away the results and then pay to buy them back. There is discussion of the future of scholarly communication and the publish or perish syndrome and a feeling that we should move away from the present system, which seems to reward fairly intransigent publishers at our expense. Nor are these the cries of fringe lunatics, but of level headed academic managers who find it impossible to justify the present system and of major library groups such as ARL.

Next in the chain come the purchasers, principally libraries. They too are uncertain of their futures. On the one hand they continue to pay complainingly for print on paper journals, while trying to corner the market in electronic services in order to ensure their role, an ambition

about as plausible as that of the Bunker brothers trying to corner the market in silver.

Finally come the consumers, the library users, their heads filled with nostalgic notions of browsing collections and a changeless world, of free libraries and quiet stacks, where there are no missing issues and volumes are miraculously bound while they sleep. Yet there is a huge mushrooming of electronic bulletin boards . Already over 5000 such boards exist on the Internet, where tens of thousands of presumably the same people communicate and debate electronically, becoming the producers who began this circle.

At the moment it is very difficult to see any basis for rational discussion and I begin to wonder whether this may not turn into a non-issue as the electronic revolution simply bypasses the journal.

The inexorable growth in end-user activity has been well documented by Professor Jack Meadows of Loughborough University. For example he has shown that between 1985 and 1991, the percentage of scientists and engineers using electronic mail grew from 38% to 70%, while, in the same period the number accessing academic bulletin boards grew from 2% to 23% and those undertaking on-line searching from 24% to 59%. (4) It seems clear to me that an electronic communication culture is developing quietly around us, or as Elbert Hubbard put it "In these days, a man who says a thing cannot be done is quite apt to be interrupted by some idiot doing it."

Curiously, the only people that we hear nothing from in this debate are the serials agents. Curious because they are perhaps the group most at risk in all of this.

Control

Let me now turn to issues of control. UK librarians have been prominent in the growth and development of networks and indeed a librarian now chairs the JANET National User Group. Major network resources such as BUBL are managed by librarians. Training projects such as JUPITER and the Mailbase project at Newcastle have again been significant for libraries. UKOLN has recently produced its statement on the importance for us all of networks. In sum we have begun to create an involvement and a cadre

of workers in the area. However there remains a significant training need in networking, one which the Follett Committee reviewing Higher Education Libraries may address. Whether or not this happens, each of us has to decide whether this is a role we want to have or one we are content to leave to others. There will be problems over job descriptions and demarcation and a tendency to start turf wars. Many of the existing network systems and products are at a development stage where they might be described as user hostile rather than user friendly. What we then have to move towards is the computer service concept of user support at our first line activity.

Having suggested that we have achieved a powerful position in the area of electronic information, albeit one that has significant training and attitude implications, I'd like to turn next to issues of national policy and control. I'm conscious that this is not a higher education conference, but it seems to me that there are some astonishing prospective changes there which will or at least could have a dramatic impact on library provision throughout the country.

Almost by chance we have acquired a datasets policy for higher education and it poses the first issue. The policy is aimed at a mass market of students and aims to provide up to twenty databases to universities over the JANET network. On present estimates this will have an annual cost of around £100,000 for each institution. That's going to knock a hole in institutional serials budgets if the cost is met entirely by the library. And it is going ahead. The new Joint Information Systems Committee has inherited a couple of datacentres and a few databases, notably those from ISI. ISI is less than delighted at the very heavy use made of the service and the impact on their hard copy sales. So we have developed the doughnut strategy. This surrounds the ISI datasets with a variety of tools, services and datasets. We shall have an ETOC - an electronic table of contents, a number of subject databases such as Embase and Compendex and a service such as OCLC's FirstSearch. The aim is, within twelve months to have something of value for everyone in higher education. A budget of several millions of pounds is set aside for this. If we can renegotiate a contract with ISI on acceptable terms, that will provide welcome jam for the

centre of the doughnut. If we fail, the doughnut still exists.

The JISC itself has a portfolio of activities, ranging from the product and training activities of NISP, through to the critically important catalytic centre of UKOLN and on to data and gateway services such as NISS and HENSA. It has also identified a new budget line to support projects ranging from navigation tools to SGML. Again some millions of pounds are involved. If you don't understand the acronyms, at this stage it doesn't matter, the key point being that there is a huge investment in electronic information services and products at national level.

Then there is the Follett Review of Higher Education Libraries. It has already identified a first budget line for improving automated services in libraries, again of some millions of pounds. It is also clear that at least some members of the review see an important need for retraining of library staff into a new (or is it an old) role as supporting teaching. It is now almost a truism to say that the increased student numbers will lead to a shift from teaching to learning and a shift to student centred learning which will tend to be library based. So there are two emerging concerns, training us in network use and teaching us to train others.

Availability

Let me turn thirdly from control to availability. We have been familiar with the commercially available services for many years, whether indexes like Medline, or full text like LEXIS. But a whole new range of not-for-profit and free or cheap resources has mushroomed in the last year or two. The network has become a vast emporium, an Aladdin's Cave. All the obvious old favourites are there, the indexes and abstracts and newer ones such as OPACs, there is full text from the Bible to Lewis Carroll. Newer forms such as census or satellite data and graphical images are there. Harvard has the world's 2000 worst lawyer jokes (Q. How do you get a lawyer down from a tree? A. Cut the rope!). Public domain software and shareware are there, from serious stuff like word-processors in Lancaster University to Part 7 of Commander Keen from Sweden. Perhaps the best guide, at least to the Internet, is the new book by Krol, *The Whole*

Internet Guide, which as well as describing networks, provides a listing of subject areas which are available.

Training of academic staff in these skills is an important issue and not an easy one to resolve. Part of the answer is generational and it will slowly be recognised that some of the most important research skills are those of information management. Part is elitism, which will not disappear. Until I was about thirty I thought that all heads of department and even more eminent beings were attractive women aged between 18 and 25. This was simply because most heads of department don't visit the library but send their secretaries and technicians. They are insulated, from the pressures which fall on more junior staff and we shall have quite a struggle to convince them that shifting resource into areas such as training and away from departmental budgets will be beneficial.

Like all truisms, the information explosion is often underregarded. Some estimates have suggested that the number of scientists in the world, and by extension the amount of research is growing by about 12% each year. The sheer quantity of information causes panic and yet the basic issues of its quality, its management and its cost remain the same. Information management tools can contain the flow of information, while various kinds of relevance measures can at least help on the issue of quality. The cost of information is perhaps more of a worry because we have not yet found adequate mechanisms to deal with the volume of information. Publishers are struggling with high development costs for electronic media, with no certainty that the new methods of scholarly communication have a place for them, while higher education is struggling with limited resources which allow it to collect an ever diminishing proportion of the research output of the world.

Training is as important for research staff as undergraduates. As new information sources are developed and better networking allows new opportunities for using and presenting information there will be a continuing need for what is unfortunately if graphically called retooling of staff.

Resource implications are as always obvious. How they are to be handled, whether by centralised, top-sliced or devolved funding

mechanisms is much less clear. However, I hope that next week the Joint Information Systems Committee will approve a major budget line of some hundred thousands of pounds for projects to work on Networked Information Retrieval Tools. [NB: this was duly agreed]. The Follett Committee is also expected to fund some work on the electronic journal. The arrival of the electronic journal has been more often predicted than the second coming, with just as many varied views of what it would be. In the narrow sense it has arrived in that there are several dozen properly referred electronic journals, although asking anyone to name them is a bit like playing that party game of Name Twelve Famous Belgians. The beginnings of the debate are usually traced to F.W. Lancaster's classic 1978 treatise *Toward Paperless Information Systems.* (6). In the fifteen years since then we seem not to have moved very far and can still be thought of as in the experimental arena. Even if there is a huge burgeoning of electronic journals, it is unlikely that they will appear in sufficiently large numbers to cause hardprint to diappear say before the end of the century.

In any case, the economics seem unattractive. SEPSU was asked to carry out a study of the electronic journal for the Follett Review, including discussions with the learned societies. On an admittedly limited sample, the clear message was that we would face parallel publication initially in printed and electronic form; if we ever moved to electronic publication only, publishers' costs would not fall; finally and in any case, publishers, including the learned societies wanted to keep electronic journal prices up as a continuing cash cow to support other activities.

When we look at networked electronic journals, the publishers demand a level playing field, presumably meaning that they want the community to carry out some of the risk associated with experimental work, so that they can concentrate only on lucrative winners. Of course, it may not mean that. "Level playing field" is one of those curious phrases that suddenly crops up everywhere without any consensus as to its meaning.

If we are to see a new era emerging in the electronic journal field it will not, I suspect, be for economic reasons, but because of changes in the pattern of scholarly communication; because what

was appropriate for the seventeenth century is inappropriate for the twenty-first century. Researchers have a relatively fixed amount of time for information gathering and a relatively fixed amount of time for research and analysis. That has not changed and will not change although the growth of knowledge might seem to demand it.

It was said of Thomas Young, the 19th century physician that he was the last man who knew everything. That is no longer practical in even a single discipline. Take the case of Physics, where the literature has grown steadily from roughly 80,000 articles in 1969 to 150,000 articles in 1991 and *Physics Abstracts* is now known as the green slime. Just to read all the abstracts would require one to read one hundred abstracts an hour. To cover a single core journal such as *Journal of Applied Physics* requires a researcher to read two articles an hour every working day. This is clearly impossible. Navigation and filtering tools are then being developed which allow two options. The first is to capture the same data as at present but more quickly, thus releasing time for the real work of research. Secondly, the same tools can be used not necessarily to minimise the time spent in information gathering, but to ensure that the collection of information is richer and more comprehensive.

Document delivery is another area of access where there is perceived to be a need for more activity. At Boston Spa we have a document supply service which is the envy of the world. There have been many experiments on new electronic document delivery systems, but the dependence on the present arrangements has perhaps led to a reluctance to innovate. As a general rule, decentralised systems will drive out centralised ones and we are clearly on the verge of a series of moves towards such decentralisation. This may come through the arrival of commercial players; it may come through the changing economics of regional co-operation and the encouragement of such local schemes, or it may come through major library groups such as CURL choosing to move into the field. Perhaps more importantly we can expect end-user activity which will bypass libraries and be paid for by credit cards. We already know that the credit card companies are keen on this since they do not expect a significant crimewave to develop around the theft of photocopies. In terms of the

availability of information, documents ordered electronically but supplied on paper remains the next step.

The last area of difficulty I should mention, albeit briefly, is that of standards. You will have heard of Z39.50 and SR and of the so-called protocol wars between OSI and TCP/IP. I don't pretend to understand very much of the intricacies of this. However I do know that if we are to have truly international networking it is self-evidently essential that computer applications can work with each other. There is a real need for the library community to be more involved in standards work. It can be tedious and nit-picking, but the profession that brought you AACR2 and faceted classification should not be put off by nit-picking. Standards work will be funded; that is not the issue, but there is a real need to find willing and involved librarians to take up some of this work.

Conclusion

It is my perception that there is not a problem over access to networked information. It is freely available. The problem here is in letting everyone know how to acquire access. In terms of control, we as a profession are in a healthy position to adopt a major role. There are substantial issues over training and retraining of library staff, over defining our role in teaching others the skills of information management and information literacy and problems over copyright. Availability is not the issue, but superfluity is. There is too much information and with conscious echoes of Ranganathan our task must be to create the tools which will get the right information to the right reader at the right time. This is not a problem of resource, at least at national level, but one of creating what we want. For the first time in my professional career, government or at least the Higher Education Funding Councils are challenging us not to define the problems, but are offering to pay for the solutions. They may be, indeed are, doing this for the wrong reasons, but the onus is firmly on us to respond.

If I try then to pull all of this together it might be summed up by a line from one of my favourite characters, Mae West. Memorably she once said "I used to be Snow White, but I drifted". That neatly sums up the issue for me. We can carry on

in the road of purity and virtue with our traditional concerns, or we can drift into this new world with its new needs and challenges.

References

1. Stone, Peter   *JANET and its applications for information and library services* (Follett Committee unpublished paper, 1993)
2. Grossman, Wendy M. *Into the Internet* (*Personal Computer World*, April 1993, pp388-394)
3. Quoted in:
   Foster, Jill *Navigational Tools* (Follett Committee Paper, [in press])
4. Meadows, A. J. *Training and education in IT* (ITs NEWS Issue 25, 1992, pp12-17)
5. Krol, Ed *The Whole Internet User's Guide and Catalog* Sebastopol CA, O'Reilly, 1992
6. Lancaster, F.W, *Toward paperless information systems,* London, Academic Press, 1978