

The Impact of Electronic Publishing on the Academic Community

Session 5: Digital libraries and archiving of electronic information

The development of digital libraries

Jack Meadows

Department of Information and Library Studies, Loughborough University,
Loughborough, Leicestershire LE11 3TU, U.K., a.j.meadows@lboro.ac.uk

©Portland Press Ltd., 1997.

[Copyright Information](#)

Requirements for a digital library

The first essential is to define a digital library, since technical questions can only be sensibly explored in the context of an accepted specification. Unfortunately, current discussions seem to employ a variety of definitions. These include a range of descriptors (such as 'virtual' or 'electronic') that can be substituted for 'digital'. Examining the variety of usage suggests that the vague part of the title is actually the word 'library'. In what sense can ways of handling electronic information be considered analogous with a library of printed publications [1]?

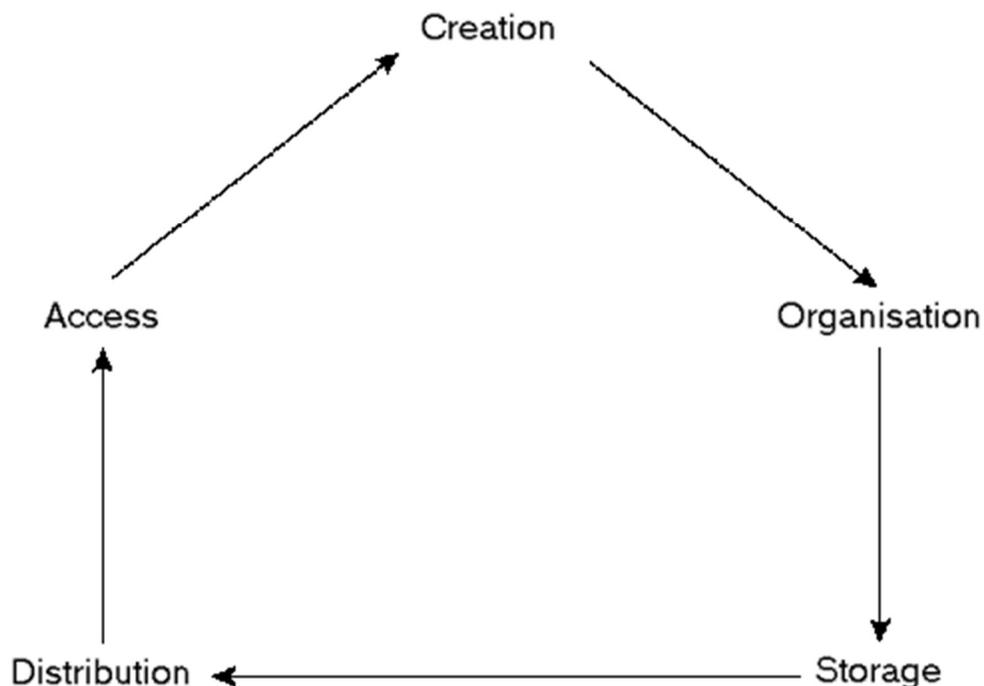
One obvious difference is that a traditional library is a physical place, whereas its intended digital equivalent has been labelled a 'library without walls'. Is it possible to go from geographically concentrated to geographically dispersed information and still apply the same label to both? Another way of putting this is --- can an online site display the characteristics we expect of a library? At one electronic extreme, it is, in principle, possible to establish a virtual reality representation of a library. A user might wander around such a library, selecting publications that look interesting, browsing through them, and even borrowing them to take 'home'. This does sound like a digital 'library', but it raises some major technical problems. At the other extreme, the alleged electronic library may simply be a focus for jumping to other services. Though this is technically much easier, the resemblance to a typical library may be small.

In essence, the question is this. Supposing we list the essential requirements of a library, how many of these can be readily fulfilled electronically? For example, if readers have queries in an ordinary library, they expect to be able to approach a librarian for help. Will electronic libraries

have similar human help available? If not, will automated help systems really be an adequate substitute? Similarly, the publications in a library have been carefully placed in a particular order so that readers can find what they want. The publications are looked after, so that they remain available for consultations by future readers. The librarians try to ensure that users of these publications obey the copyright regulations. Electronic libraries will need to match all these requirements, and all of them raise technical questions.

Looked at in another way, a print-based library is a place where a series of routine information-handling activities are carried out (along with more specialized roles). A consideration of the way that information flows occur --- especially, though not solely, in an academic environment -- suggests that libraries can be involved in all stages other than the act of creation. (See the diagram of the information cycle, [Figure 1](#).) These stages typically relate to particular types of routine activity. Thus a list of library services might include [2]: (i) the routine acquisition of information artefacts (e.g. issues of journals) in anticipation of reader demand; (ii) the archiving of such acquired artefacts; (iii) interlibrary loan (i.e. the acquisition of copies of items not held by the library); and (iv) the creation of tools which facilitate the identification and location of items held by the library.

Figure 1. The information cycle



Hybrid libraries

It might reasonably be argued that all this misses the point: the essential question is not how well an electronic library can duplicate a paper-based library? It is rather how well an electronic system can service the information needs of its customers. In other words, given the information cycle proposed above, how can it best be serviced by electronic means? But this overlooks a vital point. For some decades to come, many readers will require access to both paper-based and electronic documents. They will expect these to be available in complementary ways, which implies that there will need to be some identity of purpose between existing libraries and electronic libraries. Hybrid libraries will be common. Consequently, creators of electronic publications must take some account of the organization and structure of existing libraries. Equally, these libraries will find it necessary to modify their habits. For example, books and journals are one-way providers of information. Readers obtain information from publications: they do not expect to provide feedback to publishers and authors. With electronic transmission feedback is possible, and seems increasingly likely to occur. Can traditional libraries handle this?

Two points can be made here. The first is that much of the traditional methodology developed for providing metadata (such as cataloguing procedures) for print-based libraries can be transferred to the handling of digital information. The problem here is less one of approach than of the volume and diversity of electronic information. This will increasingly entail the provision of such metadata automatically. Secondly, print and electronic sources do, at present, have a genuine element of complementarity. For example, a traditional library contains a large number of items that are rarely, if ever, used, but finding and reading relevant material in such a library is usually not too difficult. Use of a digital library bypasses large amounts of irrelevant information and potentially taps far more relevant information. However, the relevant material is not always easy to find or to handle. Such factors suggest that a *modus vivendi* can be established between the requirements of print libraries and digital libraries.

Nor is the need for hybrid libraries the only factor suggesting there will be a role for libraries in the digital future. Another question relates to payment. The cost of electronic publications to the user has been somewhat obscured previously by the number of 'free' publications available online. Priced electronic publications are becoming increasingly common; so the question of who pays can no longer be ignored. To put it another way, much electronic publishing via the Internet has been small-scale. Larger, commercial organizations are now moving in, and will naturally expect a financial return for their efforts. Though some electronic publications will be available cheaply, others will be too expensive for individual purchasers. These are likely to be purchased institutionally and redistributed to readers. The obvious part of the institution to use for this function is the existing library. Consequently, the information chain for printed publications (author, publisher, library, reader) is likely to re-establish itself, at least in part, for electronic publications [3].

Many kinds of print-based libraries exist to satisfy differing types of demand for information. Some information niches may be filled more readily by purely electronic means than others. For example, a current German project, MeDoc (multimedia electronic documents), has brought together a range of publishers, librarians and others to investigate the setting up of a distributed digital library covering computer science [4]. This is an area where purely digital provision may prove satisfactory fairly early on. As a part of investigating its feasibility, the participants are exploring such matters as security and methods of paying. The results from specialist studies

such as this should provide better insight into the management processes required in a hybrid library.

The problem of change

Looked at from the angle of technical questions, standardization, etc., the major immediate problem for electronic publishing --- and so for libraries and readers --- is the rapidity of change. One of the great advantages of electronic publications from a reader's viewpoint is that they can be accessed from the desktop. It is no longer necessary to trek to the library only to find that the publication required is already being used by another reader. Continually changing hardware and software can affect this positive perception. Thus one consequence of change is that users must constantly re-learn the process of accessing and reading electronic publications. For example, readers were just becoming acquainted with gopher-based systems when they found they had to access Web-based systems instead. A similar irritation is that the actual on-screen image depends on the software employed, so that the appearance of an electronic publication can depend on the terminal used.

A particular problem is that existing information technology is only just able to handle the requirements of electronic publishing. Text now offers few problems (though the same is not necessarily true of hypertext), but graphical handling still has its limitations. To obtain the best results, publishers therefore aim at state-of-the-art hardware and software. Unfortunately, their audience has some difficulty in keeping up. An average staff member of even a well-endowed university is unlikely to acquire a new computer more than once every three years. Hence, a library may acquire an electronic publication but be unable to transmit it to all readers because they do not have the appropriate hardware/software to receive it.

More fundamentally, there is the question of network connections. The speed of data acquisition is obviously dependent on the slowest physical link in the network that connects the reader to the electronic publication. In addition, increasing traffic congestion is lengthening the time required to access and download an electronic publication. Reading electronic publications can therefore become quite tedious. No doubt, methods will be developed to improve the situation, but countering the overload may not be straightforward. For example, more information on the Web means more searching for relevant material. In consequence, the process of searching adds to the traffic on the system, and itself slows down retrieval.

Continuing change is, of course, also a major problem for the process of establishing standards. For electronic publishing, what is required is a high level of inter-operability [5]. That is to say, a user should be able to progress seamlessly across computer systems and services in the search for information. Inter-operability functions at a number of levels. At present, it often simply means that the various systems involved will work together in real time. Portability --- the ability of software to work with different systems --- is still relatively limited. The next stage --- data exchange --- is the one that impinges most immediately on users. Here, too, there is considerable room for progress. Typically, human input is necessary to allow the activity to function. For example, searching for information across systems is now possible, though its value depends greatly on the knowledge and skills of the user. However, it can usually only be done one system at a time, and it is up to the user to merge the results.

A major difficulty for both librarians and their customers is that they only have a limited influence on the development of the standards required for digital handling of information. These are currently controlled more by suppliers than users. One reason, of course, is the rapidity of change, which makes the long-established ways of agreeing on standards too slow. It has been said that establishing Web standards today tends to be the converse of the usual order for shooting a gun: fire; take aim; get ready. In consequence, it is often a case of the suppliers imposing *ad hoc* standards on consumers.

Technical requirements

The United States National Information Infrastructure programme has identified a number of critical technical challenges that need to be resolved if usage of electronic networking is to be both simple and efficient [6]. The listing of them below illustrates that all are important for the proper running of a digital library.

- (i) Network components that can handle voice, and text simultaneously, and can operate seamlessly.
- (ii) Information appliances and services that can provide access and services in a scalable, efficient and inter-operable way.
- (iii) Information access techniques that can enable efficient searches of large distributed information repositories, making the myriad of information resources understandable.
- (iv) Multimedia information technologies that can, for example, synchronize and integrate real-time delivery of voice and video, and can support search and retrieval based on image content.
- (v) Infrastructure for application development that can provide common solutions.
- (vi) Technologies that are dependable and manageable.
- (vii) Technologies that are easy to use and services that are accessible by users with widely varying skills, experiences, abilities and backgrounds.
- (viii) Inter-operability among heterogeneous systems will be required on an unprecedented scale.
- (ix) Security and privacy technologies that are easy to use and provide appropriate levels of security to suit the requirements, cost constraints and convenience of the end user.
- (x) Technologies and services that provide portability, mobility and ubiquity.

Though implementation of this list of desiderata would be extremely helpful for the operation of digital libraries, they are not, even so, sufficient in themselves. The list relates to infrastructural properties, whereas libraries are equally concerned with the information conveyed. For example, good information access techniques [item (iii) on the list] are important, but so is the requirement that the information purveyed should be of acceptable quality and should be archived for long-

term availability. Similarly, security and privacy technologies [item (ix)] can help control who accesses information, but a more fundamental question for a library concerns ownership of copyright in the information to be accessed.

From another viewpoint, this list can be analysed in terms of the information cycle. So far as readers are concerned the arrows in the cycle now need to run backwards. Thus the first question they face is whether they can access the information at all. If the technology does not allow this, its other capabilities, however good they may be, are irrelevant. Hence, for digital libraries, technical requirements such as those contained in items (ii) and (vii) on the list come first.

The point can be illustrated by the results of a project we have recently completed at Loughborough. This was intended to examine user reaction (mainly of postgraduate students) to electronic journals. Most potential readers said at the start that they liked the idea of electronic journals. They particularly welcomed the ability to access from their own desktops so that they would not need to make a journey to the library. They intended to select the articles they wanted to read in detail, and print them out locally.

The project ended recently, by which time a considerable number of readers had become disenchanted with electronic journals. One commented: "Electronic journals are fun to play with, but normal people read printed journals." This attitude was induced by the deterrents they had encountered, most of which concerned problems of access. Some of these stemmed from the information providers --- for example, the differing (and rapidly changing) interfaces provided by journal publishers --- but the single largest complaint related to delays in actually getting through to the journal, followed closely by delays in downloading and printing the articles selected for further reading.

In this project, access was mediated by the university library, and readers were offered personal assistance and guidance. Their overwhelming opinion was that accessing and reading the electronic journals would have been very difficult without such assistance by library staff. They did not believe that automated help facilities could have provided the same kind of informative feedback.

This leads to one final point about the technical requirements for a digital library. Though the requirements listed above will be hard to implement for networks as a whole, they are appreciably easier to implement at the level of the local area network. What is happening, therefore, is that libraries are increasingly trying to provide a digital environment within their own institutions. This bottom-up approach makes sense under current circumstances, but it inevitably raises questions regarding standard practices between different institutions, and their compatibility. De Montfort University in the U.K. has been developing a digital library environment for some years past. The lessons they claim to have learnt (as at the end of 1996) might be summarized under the following headings [7].

(i) The digital library will develop more quickly than people think.

(ii) The digital library is still a complex, unstable entity for which little theoretical structure exists.

- (iii) Because of this inherent instability, investment and implementation are still high risk.
- (iv) Digital libraries operate in a global environment: new products and services can become *de facto* standards very quickly.
- (v) Co-operation is therefore a key factor for maintaining competitiveness.
- (vi) The content of the digital library will become the dominant factor.
- (vii) Copyright issues will be resolved or side-stepped because the market will demand it.
- (viii) The economics of the digital library are not yet well understood.
- (ix) Library jobs and roles will change very rapidly.

Conclusion

The current problems of electronic access and use take us back to the first question concerning the nature and purpose of an electronic library. Various points have emerged from our examination as requiring further discussion. For example: (i) for the foreseeable future, many readers are likely to require information from a mix of printed and electronic sources. Hence, one of the key questions to be asked immediately is how best to provide this mix in such a way as to give maximum assistance to readers. (ii) If the technical sophistication of information handling continues to develop rapidly, how can a library (physical or electronic) help users who have widely varying abilities in the handling of digital information? (iii) How is the long-term storage of electronic publications to be handled, and how do questions of copyright, etc., affect the possible technical solutions?

What these kinds of queries illustrate is that the development of the digital library will not only depend on developments in technology and in the provision of information. A whole range of social, organizational and economic questions also need to be resolved, on a fairly short time-scale, if digital libraries are to make rapid progress.

Notes and references

1. A useful analysis of the digital library concept has been provided by Harter, S.P. (1997) *Scholarly communication and the digital library* (self-published; e-mail address: harter@indiana.edu)
2. Lester, R. (1997) The need to add value. In *Towards a worldwide library: a ten year forecast* (Helal, A.H. and Weiss, J.W., eds.), pp. 13--31, Essen University Library Publication no. 21

3. The following collection of papers provides insight into various aspects of the library and electronic publications: Halal, A.H. and Weiss, J.W. (eds.) (1996) Electronic documents and information: from preservation to access. Essen University Library Publication no. 20
4. Boles, D., Dreger, M. and Grossjohan, K. (1996) MeDoc information broker --- harnessing the information in literature and full text databases. <ftp://ls6.informatik.uni-dortmund.de/pub/doc/publications/nir96/boles.ps.gz>
5. Some interesting comments on inter-operability occur in: Borgman, C.L. (1997) From acting locally to thinking globally. *Library Quarterly*, July 1997
6. See the proceedings of the workshop on research and information needs available from EDUCOM at nii-forum@educom.edu
7. For details, see: Collier, M.W. (1997) A model for the electronic university library. In *Towards a worldwide library: a ten year forecast* (Halal, A.H. and Weiss, J.W., eds), pp. 180--190, Essen University Library Publication no. 21

©Portland Press Ltd., 1997.

[Copyright Information](#)