5

The Electronic Library

THE ELECTRONIC LIBRARY
ELECTRONIC DOCUMENTS
THE INCREASE OF ELECTRONIC DOCUMENTS
PAPER IS BEST EXCEPT ...
REINVENTING THE LIBRARY
PRIOR EXPERIENCE
BIBLIOGRAPHIC ACCESS IN THE ELECTRONIC LIBRARY
THE ARCHITECTURE OF THE ELECTRONIC LIBRARY
COEXISTENCE OF PAPER AND ELECTRONIC DOCUMENTS

THE ELECTRONIC LIBRARY

We use the term "Electronic Library" to describe the situation in which documents are stored in electronic form, rather than on paper or other localized media. Note that paper copies of electronic documents, or of excerpts from them, can generally be produced for the reader's convenience. However, the essence of the Electronic Library is that documents are stored and can be used in electronic (or similarly machine-readable) form. In this discussion the term "digital" is used to mean text in which individual letters, numbers, and other characters are each separately coded and the term "digitized" is used to denote an image composed of lots of bits. One could well have a digitized image of a page of text, but in that case none of the individual characters would be coded as characters. In some cases where the characters are regularly formed, it may be feasible to use pattern recognition techniques to identify the individual characters within the digitized image and, thereby, to derive a digital version of the text from a digitized image of the text. A digital record is more economical and, ordinarily, more useful.

The adoption of computers for libraries' technical operations, the transition from the Paper Library to the Automated Library, can be viewed as an evolutionary development. Much of the change represented, at least initially, the mechanization of previously manual procedures of the Paper Library. The changes have been, at least until the provision of on-line catalogs, mainly for internal efficiency and for the convenience of library employees. In contrast, the rise of the Electronic Library, in which materials are stored in electronic form, may seem more revolutionary than evolutionary because of the implications for the provision and use of library services. But is it really so radical a change? Where are the impacts on the provision of library service? How are we to achieve a graceful and efficient continuity of service as electronic documents come into use?

Much of what is published nowadays has already existed in machine-readable form because authors commonly use word-processors and printers often employ computer typesetting. Frederick Kilgour has characterized the evolution of electronic publishing as being in four stages:

1. Preparation of text in machine-readable form for the production of paper copies with subsequent discard of the machine-readable version,
2. Publication of paper copies and also in machine-readable form,
3. Publication in machine-readable form only, and
4. Publication of information in machine-readable form in an encyclopedia-like database. (Endnote 1).

Nevertheless, the "full text" of documents in machine-readable form has been generally absent from library services until recently, in contrast with the progress made in making bibliographies, catalogs, and numerical data available in electronic form.

ELECTRONIC DOCUMENTS

The characteristics of electronic documents differ greatly from those of paper documents, which were discussed in Chapter 2:

1. Electronic documents are not localized. Given telecommunications connections, an electronic document can be used from anywhere, without one even knowing where it is stored geographically.

2. In practice several people can use the same database or electronic records at the same time.

3. Electronic documents are easily copied.

4. Documents stored electronically are very flexible. They are easy to revise, rearrange, reformat, and combine with other documents. Hence the popularity of word-processing among people who have to create and, more especially, revise documents.

5. Collections of documents stored in electronic form are now less bulky than paper versions. The trend is to even greater compactness.

In each of these five important characteristics, electronic documents are quite different from paper documents.

THE INCREASE OF ELECTRONIC DOCUMENTS

There is a steady growth of documents in electronic form and access will necessarily have to be provided to them. Databases are increasingly available in (and sometimes only in) machine-readable form. Publishers commonly have or could have the text of their books in machine readable form even if they may not yet choose to publish them in that form.

The most obvious source of electronic documents is new publications issued in electronic form. But what of the older materials on paper that occupy so many miles of libraries' shelves? Libraries have undertaken a major, systematic effort at the retrospective conversion of older catalog records from cards to electronic records. What of the retrospective conversion of the texts of older paper documents themselves? The idea might seem wildly unrealistic, but there are grounds to believe that, over time, significant and increasing amounts of older material will become available as electronic documents. In selected areas, notably literature, texts have been converted for research purposes: All classical Greek texts and increasing quantities of medieval and modern literary texts are already available in electronic form. Devices have been available for some years that can scan printed material, derive digital versions, and "read" the text out loud for the blind and visually-impaired. The same approach can be used to convert paper texts into electronic form as an alternative to keying them when an electronic form of the text is needed for word processing purposes. These electronic copies are usually discarded or, at least, are not made systematically available. They could be.

The latest fax technology, using the CCITT class IV standards, involves the transmission of pages of documents by means of digitized images of pages. Interlibrary loans sent by CCITT class IV fax could be stored and, probably, whenever feasible, converted to digital form (i.e. coded as individual characters rather than images of text) for ease of use and economy of storage. Technology exists for copying microfilmed materials into electronic, digitized form and, most likely, much of this, once digitized, could be also be converted to digital text. Therefore, it is technically feasible that very substantial amounts of older as well as future library material becoming available in electronic form whether or not librarians engage in retrospective conversion. Storage costs for electronic documents are decreasing steadily, while the building costs for storing paper documents are not.

PAPER IS BEST EXCEPT ...

A document on paper, such as a letter or a book, is unquestionably extremely convenient to use compared with other media such as microfilm or floppy disks, at least for most purposes. But, even if we were firmly agreed that paper were the best medium for documents, it is increasingly clear that there are significant exceptions to this rule. Electronic documents, with or without the generation of paper copies, become preferable:

1. When documents are highly volatile. For example it is unwise to depend on printed paper versions of highly changeable material such as airplane schedules, stock prices, and currency exchange rates.

2. When *manipulation* of the document is desired. No one would want to have to transcribe (and possibly mis-transcribe) printed numerical data for statistical analysis if the data were already available in electronic form. Similarly, when a text is to be modified, a bibliography revised, or the layout rearranged, the availability of an electronic copy in a standard format for word processing can dramatically reduce the work involved compared with having only a typed or handwritten copy.

3. When *scanning* for names or for particular words or phrases in a lengthy document. Trying to find mention of some thing or person in, say, a multi-volume printed work or a run of periodical is very tedious and error-prone. No one now would want to compile a concordance "by hand" anymore: The first step in concordance making now is the creation of an electronic copy of the text.

4. When light use of *remote material* is needed. For a thorough reading of a document that is not available locally, obtaining a paper copy by interlibrary loan would probably be preferred. If, however, use were light--to check here and there in the document or to skim the document superficially to see whether or not a more careful reading would be warranted--rapid access by telecommunications to an electronic copy could be attractive.

5. When rapid *communication* is desired, especially within a dispersed group that is not conveniently available at the same time and place, the use of electronic mail has considerable advantages over ordinary mail and, for some purposes, over telephone.

Note that these examples do not include the usual notion of solid, systematic, consecutive reading. They could be regarded as exceptional cases around the fringes of "normal" use of documents, but in some circumstances, as when geographically separated quantitative researchers collaborate, these exceptions could add up to substantial amounts of activity and a significant proportion of total use. Electronic documents add new possibilities for the use of texts and, in this regard, constitute an enhancement that is valuable in its own right.

REINVENTING THE LIBRARY

What are we to with a document in electronic form? There is little choice but to do the same as we do with a paper document or with microfilm document:

Catalog it and, as with manuscripts, pay careful attention to which version or state of text it is.
Store it in some accessible place.
Give it a call number.
Ensure that pertinent bibliographic and location data are accessible in or through bibliographic databases.

There seems no real alternative. Given that electronic documents exist and are becoming progressively more important, to ignore them would be to provide a progressively less complete library service. A library administration might choose to retain an exclusive concentration on paper, microfilm, and other localized media, but that would mean that access to electronic documents would have to be found through other channels, such as the computer center. The result would be a split in the provision of library service: the "library" providing access to only some kinds of documents; and another organization providing the balance of the library service--that which involves access to electronic documents.

The significant difference with an electronic document is that if you have the call number it should in principle be possible, from any workstation, to gain access to it remotely, view it, download it, and, in brief, "use" it. Think how much simpler and quicker it would be if librarians and, even better, library users could obtain their own interlibrary "loans" (now, technically, copies or excerpts) on a self-service basis, requiring the tolerance but not the time or energy of the staff of the library from which it is obtained. This change would be rather like the change from having closed library stacks, in which library employees had to fetch each book for users, to open stacks in which library users could obtain and examine books by themselves. Similarly, in the Electronic Library, library staff would be mainly concerned with creating and sustaining the system so that users could serve themselves.

Self-service, however, is a mixed blessing. It also assumes standardized, intelligible procedures, presupposes some expertise on the users' part, and may make it less easy for the service providers to know what is going well and what is not going well. Yet it may be the only affordable way to support large-scale library use.

PRIOR EXPERIENCE

Fortunately we already have some familiarity with what to expect in the Electronic Library from our experience with the transition from the Paper Library to the Automated Library and from our experience with on-line bibliographies. The change from the Automated Library to the Electronic Library is an extension of the same changes that have characterized the shift from the Paper Library to the Automated Library (as noted in Chapter 3 above):

standardization of data,
remote access to files,
the linking and combining of files,
access to numerous different files from the same terminal,
increased cooperative use of shared files,
discontinuation of numerous, more-or-less duplicative local files,
greater capability for doing things to and with the (computer-based) files, and
increased vulnerability to technological failure.

Hardly recognized is the fact that many librarians and library users, particularly in academic and special libraries, already do have extensive experience with the Electronic Library through their use of on-line bibliographies. This is because bibliographies tend to occupy an ambivalent position: They are acquired and used primarily to provide access to other documents, but a bibliography is itself a kind of document. Bibliographies added to library collections are "library materials." Over twenty years there has been a steady shift towards searching bibliographies on-line as well as or instead of on paper. Although not usually viewed as such, this is an early, incomplete manifestation of the Electronic Library with which there is already substantial experience and familiarity. Hence we should expect the future expansion of the Electronic Library to be an expansion of what is already familiar.

BIBLIOGRAPHIC ACCESS IN THE ELECTRONIC LIBRARY

The advent of text in electronic form, the step from the Automated Library to the Electronic Library, has two profound consequences for bibliographic access and catalog design. First, card catalogs are, necessarily, physically separate from the physical (paper) documents that they describe. Given the technology of paper and of cardboard, it could not be otherwise. Further, *use* of the documents involves physical handling, often borrowing. However, to the extent that both catalog and text are in machine-readable form, both would be remotely accessed from the same workstation and the former physical separation between catalog and text becomes unnecessary–or, at least, transparent and irrelevant to the library user, who should be able to move effortlessly between catalog and text. The catalog might be searched and one or more catalog records retrieved. Then the user might want to examine the contents of a book. A book on paper is more than a mere string of characters since there is an extensive internal structure of references: from the table of contents to chapters and sometimes sections within chapters; form the index entries to numerous points in the text; and, often, internally from one part of the text to another. (Texts in electronic form with this connective apparatus are known as "hypertext". Here, again, the basic ideas were anticipated early in the century by Paul Otlet and his "monographic principle", but the constraints of the technology of cardboard, loose-leaf binders, cutting and pasting simply do not lend themselves to effective hypertext).

In an on-line world the user could move to the table of contents by depressing a key, then on to examine a chapter. Next the user might want to look for specific terms or names in the index, on-line, then move to specific patches of text, again on-line. Since the text is on-line one could expect a concordance providing access to all of the text. The user might abandon that text, follow up a reference (from inside the text or from a citation index) to another text, go back to the catalog records to look for another book, or scan the subject headings with a view to reformulating the search. There would be a continual changing, "zooming in" and out between a broad view and focus on details. It is not that the familiar data elements of the catalog record will have disappeared or that the identifying and locating functions are any less important, but rather that the catalog will effectively have disappeared as a recognizably separate, physical entity. Instead, the catalog data would be part of a much broader set of data elements and the catalog function would have become one feature in a suite of related functions in on-line library use.

The second consequence for bibliographic access and catalog design in an Electronic Library is that the traditional justification for having a catalog begins to disappear. Historically a library catalog was a guide to local *holdings*. Yet for a library user what matters is convenient *access* to texts. With documents on paper, what is locally owned is (approximately and imperfectly) what is conveniently accessible. In practice, studies have shown that, at least in university libraries, users typically cannot find 40 percent of locally owned material on the shelves when they seek it. However, with convenient telecommunications, the physical location of an electronic text is substantially irrelevant. Databases (which are copied, not borrowed) at a distance could be more reliably accessible than paper documents owned by one's local library. What is needed, then, is a bibliography, or union catalog, of what is conveniently accessible rather than the much narrower concept of a catalog of what happens to be locally stored.

Three elements–documents, bibliography, and holdings records–remain the needed elements: Documents becoming available in electronic form will need to be stored somewhere; bibliographies will continue to be published; libraries, as documents stores, have to have their copy-specific inventory and status records. In the previous chapter we concluded that since bibliographies constitute the principal means of identification, there should be extensive, convenient access to bibliographies regardless of technology; and, because it is necessary not only to identify but also to locate material, it should be made possible, when searching bibliographies, to ascertain the call-number and availability of documents.

Because of the nature of paper, in the Paper Library and in the Automated Library, *conveniently accessible* meant, for practical, physical reasons, "locally held." Given some major assumptions about telecommunications and adherence to standards, the close coupling of "conveniently accessible" and "locally held" begins to dissolve with the Electronic Library. Because electronic documents are remotely accessible, it does not matter much to the library user where the documents are located--any more than it matters much to the user of an automated library where the disk drives of the on-line catalog are located. There are economic, engineering, and security considerations concerning the storage of electronic records but these are technical matters for those who provide the service and of limited concern to the library user. In brief, in the Electronic Library "conveniently accessible" ceases to mean locally held. It becomes as foolish to want to limit library users to locally held documents in the Electronic Library as it would be to want to limit cataloging in the Paper Library to documents published in even years.

In order to achieve the central mission of libraries--providing access to documents--references in bibliographies should link directly to and enable immediate reference to electronic documents as and when available. The technical and "architectural" aspects of achieving such linking is outside the scope of this book. However, the general pattern can be seen as an extension of the restructuring discussed in relation to catalogs in Chapter 4. The diagram showing a simplified record structure for the Automated Library in Fig. 4.3 is repeated in the first part of figure 5.1 and is extended to include linking to an electronic document in the second part of figure 5.1.

[BEST PLACE FOR Fig. 5.1]

There appears to be no obvious technical reason why access to on-line bibliographies and union catalogs should not extend to include the call numbers of electronic documents wherever located--just as we would want the call numbers of local holdings of paper documents in the Automated Library. Who would want to settle for less?

To the extent that texts become available in electronic form, the whole view of library collections changes. The location and ownership of copies of texts becomes a technical detail for librarians but irrelevant to the reader: What counts is what is conveniently accessible. Given modern telecommunications, any attempt to restrict users' attention to locally stored electronic documents would be a travesty of long-established traditions of library service.

THE ARCHITECTURE OF THE ELECTRONIC LIBRARY

What would it take to build an Electronic Library and, indeed, to make Electronic Library service common practice? To develop a library with electronic documents we do not appear to need to draw on anything in librarianship that is different from existing principles. Rather, as with paper and with microform, we have to interpret the same familiar principles in ways appropriate to the technical characteristics of the medium. With electronic documents, even more than with microforms, adherence to standards is important for progress. Electronic documents should themselves be in standard formats. Standards are needed for cataloging electronic documents. Communications formats are needed for conveying electronic documents. Substantial and compatible telecommunications protocols are of great importance. Much work needs to be done in developing and adopting compatible national and international standards for characters, images, documents, telecommunications, and so on. Also, of course, the concept of an Electronic Library assumes a substantial and expensive infrastructure of computing capacity, data storage, and telecommunications, which in turn, requires expertise for successful use. These investments are being made anyway in some of the contexts in which library services are being provided.

COEXISTENCE OF PAPER AND ELECTRONIC DOCUMENTS

Although the growth in electronic documents can be expected to be dramatic, the proportion of documents that are available in electronic form (relative to paper documents) will vary greatly from one situation to another. Further, as noted above, for some purposes paper documents are preferable. Not all documents will be available in electronic form. Paper documents are unlikely to disappear and it is undesirable that they should. So discussion of the Paper Library *versus* the Electronic Library is likely to be of limited benefit. Just a few basic issues are important for planning:

1. Electronic documents are becoming increasingly important and arrangements to provide access to them–the Electronic Library–*must* be developed and is best viewed as additive. The world is changing and this additional form of library service appears to be not only desirable for library users but also inevitable.

2. Library services from now on will have to provide access to paper documents and to electronic documents according to their users' needs. In other words, library planning should be based on the assumption that all libraries will evolve into some *combination* of an Automated Library and an Electronic Library. The balance between paper documents and electronic documents can be expected to vary widely between libraries and over time. What the balance is at any given place and time matters far less than ensuring that libraries are equipped to provide access to both kinds of document.

3. The difference between the Electronic Library and the Automated Library is in the technology of the documents, not in the bibliographic access. Bibliographic records need to include information indicating the physical properties of each document, (e.g. paper, microform, electronic, optical), but a unified bibliographic approach for all documents is most likely to be helpful. The bibliographic apparatus and internal procedures of the Automated Library would not need to change much for the shift from Automated Library to the Electronic Library. See figure 5.2.

[BEST PLACE FOR FIG. 5.2]

4. The view of the Electronic Library as an *addition* to the Automated Library needs to be modified in two ways. First, purely additive funding should not be assumed. Even though a plausible argument could be made that, in any given case, funding is insufficient to support the

Automated Library (as, before it, the Paper Library), nevertheless it is to be expected that service priorities will increasingly lead to a reallocation of total library resources from the Automated Library to the combination of Automated and Electronic Library. In this sense the Electronic Library may well be additive since it is a new form of service, but in its claim to resources, subtractive from the Automated Library. It should be a matter of very careful deliberation and consultation to determine when and to what extent, in any given situation, library service should change from only being an Automated Library to being a combination of Automated and Electronic Library service.

Second, in practice there will be some acceptable scope for the substitution of electronic for paper documents, of the Electronic Library for Automated Library. At some point of adequacy and reliability in on-line catalog service, it becomes reasonable to abandon the card catalog with its high maintenance costs. At some point, in at least some cases, subscription to printed bibliographies cease to be justified when on-line versions can be searched conveniently. In the same way, it is reasonable to expect that in some circumstances where there is a choice, access to electronic documents will be preferred to use of paper documents even when the paper version is conveniently available.

The key to consideration of the Electronic Library is recognition that providing access to electronic documents will be needed. How the balance between paper and electronic documents will evolve is an interesting but less urgent issue.

Endnotes for Chapter 5: The Electronic Library

1. Frederick G. Kilgour. *Beyond Bibliography*. (London: British Library, 1985).

2. For a fuller discussion of the implications of texts becoming available in electronic form see Michael K. Buckland, "Library Materials: Paper, Microform, Database," *College and Research Libraries* 49:(1988): 117-22.

3. For academic libraries, the following provide convenient introductory reviews of material available in electronic form: Margaret Johnson and others, *Computer Files and the Research Library* (Mountain View, Calif.: Research Libraries Group, 1990); Constance C. Gould, *Information Needs in the Humanities: An Assessment* (Stanford, Calif.: Research Libraries Group, 1988); Constance C. Gould and Mark Handler, *Information Needs in the Social Sciences.* (Mountain View, Calif.: Research Libraries Group, 1989); Constance C. Gould, *Information Needs in the Sciences: An Assessment* (Mountain View, Calif.: Research Libraries Group, 1991).

4. For an introductory discussion of the characteristics and acceptability of different media see Brian Shackel, "Information Exchange within the Research Community," in *Information Technology and the Research Process: Proceedings of a conference, Cranfield, 1989*, ed. Mary Feeney and Karen Merry (London: British Library Research, 1990),147-62.

5. In the terminology of hypertext, cross references (*links*) are made between passages of text (*nodes*) which can, in principle, be text, graphics, animation, or digitized sound. Two or more documents with links between them constitute a *hyperdocument*. A *hypergraph* is a map of *links*. For an introduction to hypertext and its complexity see N. M. Delisle & M. D. Schwartz, "Contexts–A

Partitioning Concept for Hypertext," *ACM Transactions on Office Information Systems* 5 (1987): 168-86.

6. Paul Otlet, *International Organisation and Dissemination of Knowledge: Selected Essays* (FID 684). (Amsterdam: Elsevier, 1990).