
Archival Issues in Network Electronic Publications

MAYNARD BRICHFORD AND WILLIAM MAHER

ABSTRACT

ARCHIVES ARE RETAINED INFORMATION systems that are developed according to professional principles to meet anticipated demands of user clienteles in the context of the changing conditions created by legal environments and electronic or digital technologies. This article addresses issues in electronic publishing, including authentication, mutability, reformatting, preservation, and standards from an archival perspective. To ensure continuing access to electronically published texts, a special emphasis is placed on policy planning in the development and implementation of electronic systems.

INTRODUCTION

Archives are information systems retained for their long-term value—or, if one is an optimist, their permanent value. Archival theory and practice are based on seven areas of professional responsibility. Archives are established, administered, and evaluated by institutions, organizations, and individuals to ensure the retention, preservation, and utilization of archival holdings. They are authenticated by analyzing their content and obtaining evidence of their source during the process of their accession. They are appraised or evaluated on the basis of their anticipated use in relation to the costs of description and long-term retention. They are arranged according to source in the original order or structure in which they were kept while in active use. They are described in inventories, finding aids, and guides

Maynard Brichford, Library, University Archives, Room 19, 1408 W. Gregory, University of Illinois, Urbana, IL 61801

William Maher, Library, University Archives, Room 19, 1408 W. Gregory, University of Illinois, Urbana, IL 61801

LIBRARY TRENDS, Vol. 43, No. 4, Spring 1995, pp. 701-12

© 1995 The Board of Trustees, University of Illinois

to facilitate long-term access to their informational content. They are preserved to assure their future availability in a safe environment and on media that will remain accessible or renewable for the period of anticipated use. They are used to explain the past, provide guidance in the present, and accountability to the future.

Archival programs are the repositories of information systems from all types of persons and organizations. They must cope with vast accumulations of recorded data, as well as languages and formats representative of all past and present recording technologies. The three principal categories of archival materials are official files of institutions and organizations, publications issued by such bodies, and personal papers of individuals. Archival holdings in each of these areas are unique in the sense that they are treated as aggregates accumulated by a specific person or corporate body. Electronic information technologies have had profound effects on aspects of all of these categories.

Large quantities of information have been generated, transmitted, received, and stored on electronic disks, tapes, and other formats. The primary archival concern with regard to electronic publishing is that the published material should be transferred to archival custody. When the transfer occurs, the archivist must address the issues of authentication, appraisal, arrangement, description, and preservation or physical protection. These responsibilities are closely interrelated. Most archival holdings are not identical copies of information obtainable in other repositories, so they are appraised, accessioned, and retained according to their potential value and the anticipated costs of processing, preservation, and retention. Authentication and anticipated use have a direct effect on appraisal. Description and preservation of the information affect both its future availability and value.

As custodians of aging information systems, archivists are aware that they maintain information, not artifacts, and that reformatting is an essential tool of their practice. Few archives have sufficient intrinsic value to justify expensive efforts to ensure the long-term retention of original formats (Mitchell, 1956, pp. 139-42; National Archives, 1978, pp. 1-6). The most effective way to satisfy archival requirements for handling electronic information is the establishment of procedures and standards to ensure that valuable material is promptly transferred to archival custody in a format which will permit access on equipment that will be readily available in the future. Few archives will have the resources to copy data from obsolete formats for use on each future generation of software and hardware. Textual, quantitative, and graphic information on paper has been the standard format for the issuance of official or formal information. Archival copies are often

published or printed out with a date and an indication that the information is a record copy. Given the markets for competing standards and revisions, scanners may provide an acceptable option for converting information on paper to future electronic formats of choice. Long-term costs and access requirements are the crucial factors in determining how much information should be retained in electronic formats.

Electronic publications include journals and periodicals, books and monographs, newsletters, bulletin boards, and open interest-based listserv communications. It excludes e-mail and personal communications between individuals. All recorded, published, or disseminated information in electronic devices has a physical existence, including:

1. a location (hard drive, diskette, or tape);
2. a structure or organizational framework (program) for data entry, storage, and retrieval;
3. variable data entered by humans or systems designed by humans;
4. telecommunications or a means of communication or distribution of the information to an extended group; and
5. transmission standards.

As information ages, the user clientele shifts from subscribers to commercially viable or subsidized networks to researchers whose primary interests are in searching older information for "leads" and in understanding the developmental processes represented by information systems. The results of scholarly research are published to meet current user interests. Many current scholarly publications are a means of mass communication and represent both the product of research activity and a source for future investigations. Noncirculating record copies of such publications are transferred to archives as evidence of the functions of the agency disseminating the information. The principal archival interest is in the long-term or secondary use of information. Such uses are more likely to relate to process or the context of events rather than specific transactions. With the passage of time, scholarly studies of archival information systems can benefit from increased detachment and perspective and the loss of substantial portions of detailed information, which is destroyed after a few years.

AUTHENTICATION

Authentication involves a determination of the validity or integrity of information. Integrity requires the unbroken custody of a body of information by a responsible authority or individual. Continuous custody is assured by providing policies and rules for changes and by

the arrangement of information so that the creating agency is identified and the dates and parties responsible for additions, modifications, and deletions of records are recorded. Publication or distribution of multiple copies by providing multiple access points for copying should also be documented to establish the authenticity of the information. Authenticity is a condition precedent to the appraisal of information systems because what will be kept depends on the credibility of information (Graham, 1994; Brichford, 1977).

Cost, mutability, and miniaturization of information affect the choice of an electronic format and often result in a medium that has a relatively short life. The original is the first recorded version of information. In the case of electronic media, it may be revised, copied, or deleted with ease. Information in such formats may be suspect due to the ease with which it may be altered. From an archival perspective, the value of information is dependent on its content and the custodial responsibility of the agency that maintains it—e.g., the source determines authenticity. The authentication of archival information requires that it be verified as to source, date, and content. The mutability of electronic network data is a common characteristic of aging information systems. Chronicles or reference data have been updated at regular intervals by adding new information. Electronic publications may be supplemented at any time provided the circumstances of the additions are stated and provided that existing data are not altered. Information that is mutable, modifiable, or changeable loses its validity if the persons adding, altering, or deleting information cannot be identified and the time, place, and nature of the changes is unknown. The long-term retention of electronic publications is also a problem because of the lack of archival standards of permanence for digital storage media.

PRESERVATION IS ACCESS

The proper consideration of the preservation of electronic texts must start by reexamining the very purpose and meaning of archival preservation because the requirements of electronic texts are fundamentally different from traditional library and archival materials. To date, most literature and professional association offerings on preservation consist of detailed descriptions of techniques and technical standards employed to increase the longevity of physical media. Without discounting the importance of these technical issues, the archivist's and librarian's preservation responsibility is better understood as a matter of ensuring the future availability and intelligibility of the informational content of documents.

In this sense, preservation is more a matter of access to information than it is a question of survival of any physical information storage media. In fact, by looking at electronic information systems and

technologies, one can see clearly that what matters most to the user is the survival of the information itself and the access points provided by the system and not the specific hardware, or even software, on which information is stored.

TRADITIONAL PRESERVATION CONSIDERATIONS

Protection of the physical media or information carrier is not entirely irrelevant, and a brief review of the threats to the physical survival of library and archival materials is a useful starting point because many of these same issues pertain to the physical longevity of electronic storage media. The conditions which influence the longevity of materials and the nature of preservation work can be categorized as internal and external. Internal conditions arise from the nature of the material on which information is recorded. For example, the longevity of paper documents is heavily dependent on papermaking processes, especially since the 1860s when lignin in wood fibers and chemicals used to process wood into paper resulted in the development of acid compounds that have caused paper to become brittle, stained, and fragile. Other media, such as photographic prints and negatives, sound recordings, videotapes and computer tapes, disks, cards, and perforated tape, present comparable problems because of the chemical instability or fragility of materials ranging from glass plates and nitrate films to acetate tapes and oxide coatings on tapes. Generally there is little that can be done about internal or inherent characteristics of the documents other than replacing the media through expensive and often imperfect methods of recopying or performing major chemical and physical treatments such as deacidification and encapsulation.

Regardless of media type, external threats are considerable. Light, especially ultraviolet, fades inks and photographic images. Heat accelerates the acidification of paper and the chemical and physical breakdown of nonpaper media. For example, excessive heat can cause the oxide coatings on which magnetic signals are recorded to separate from the backing of audio and videotapes. Excessive moisture, whether atmospheric or liquid, causes considerable damage—inks run, papers and photographs fuse together, molds and mildews grow. Atmospheric moisture also acts as a catalyst for the acidification of paper. Air pollution causes damage both through particulate matter and gaseous fumes such as sulphur dioxide and nitrogen dioxide, which are especially harmful to plastic-based materials. Insects and rodents damage documents both through their chewing and droppings. Adjacent materials pose a serious threat. Particularly harmful are newsprint clippings, paper clips, pressure-sensitive tape, rubber bands, inadequately sealed wood cabinets or shelving, and magnetic

fields, which can weaken and destroy information on tape recordings or computer diskettes. Perhaps the greatest external danger to documents is posed by humans who expose records to all of the hazards listed above during their creation, archival processing, and research use. Through carelessness, and occasionally maliciousness, humans erase, destroy, and steal documents.

To address the diversity of threats to the longevity of documentary material, archival preservation traditionally has focused on techniques to extend the longevity of the information carrier. Common preservation work has included: controlling temperature, humidity, and light in stacks; physically protecting materials in boxes, containers, or encapsulations; cleaning and removing harmful materials; neutralizing acid in paper; and educating records creators, staff, and users in safe handling techniques. For modern documentation, recopying documents onto archival-quality materials—e.g., silver halide microfilm or acid-neutral paper—has become an important aspect of preservation.

PRESERVATION OF ELECTRONIC MEDIA

In the threats to traditional documentary materials, one can see several hazards to electronic media, whether "floppy" diskettes from an author's personal computer or a mainframe's mass-storage disk packs and tapes. Applying the paradigm used for conventional archival materials, one would examine the chemical and physical composition of computer disks and tapes, the volatility of the plastics, the stability of the recording strata, and then employ accelerated aging tests to assess the effect of environmental factors such as humidity and airborne gases (see Cuddihy, 1980, pp. 558-68). After all this research, one could outline standards for "archival quality" storage media and environments. These are important activities about which archivists should be well informed, but to approach the preservation of electronic texts by focusing on physical threats will miss the far more pressing matter of ensuring continued accessibility to the information on such storage media.

A different approach to electronic texts is needed for three reasons. First, unlike conventional paper and photographic materials which remain eye-legible even in advanced states of deterioration, electronically-recorded information can only be examined and used if the hardware and software on which it was created remains available and operative. A twenty-year life for the plastic backing material used for computer tapes and disks is irrelevant if the tape or disk drives on which they were recorded become obsolete and unavailable after ten years. What remains is not the information but a long-life physical artifact that rather ironically might find its best use as a paperweight

(Stielow [1992, p. 339] cites one such example in which even the manufacturer was unable to provide the software necessary to read a five-year-old CD-ROM).

Second, the usable life of physical media for electronic information storage has progressed to the point where, with reasonable operating precautions, it is now greater than the life cycle of most software and hardware used to access the media. Certainly disks and tapes can be damaged through accident and carelessness, but the actual media are stable enough to survive until the information can be recopied and verified onto new media (for a bibliographic introduction to literature on the preservation of magnetic recordings media see Child, 1993). In addition, recopying is an option for digital information which can be copied faithfully multiple times without image degradation, unlike conventional materials, including audio and videotapes. Thus, preservation actions can and must be taken at regular intervals, far closer to the original creation and use of the electronic information.

Third, a different approach to the preservation of electronic texts is needed because of the increasing complexity of information on such systems. Merely ensuring the longevity of a computer tape and providing a means to read the files it contains will not necessarily preserve the complex nature of information linkages that are at the core of the most interesting electronic publishing ventures. Simple "dumps" of files or printouts of data, as one might do in adherence to the traditional preservation practice of reformatting, will not preserve the dynamic nature of access systems or the hypertext links in electronic publications. In such cases, the access points and linkages can be a key part of the original publication and thus must be maintained if one is to truly "preserve" the document.

By examining the relation of traditional archival preservation to electronic publications, it becomes clear that there are only three options for ensuring ongoing accessibility:

1. The texts can be off-loaded by printing onto hard copy once sufficient time has elapsed that electronic and interactive access is no longer critical.
2. The original storage media, software, and hardware can be retained to allow continued access to the system. Under this option, the publisher, library, or archives must become a hybrid of a museum and a specialized electronics laboratory.
3. The data, access systems, and hypertext linkages can be continually converted and verified as each generation of hardware and software is replaced or upgraded.

All factors being equal, the third option is preferable. However, it may be viable or appropriate only in circumstances where electronic

publishers have the incentive to provide the considerable resources required, and where there is a substantial ongoing user community interest in the information.

EDITORIAL AND ADMINISTRATIVE POLICY ISSUES

Implementation of any of these options illustrates the fundamental difference and challenge of electronic information systems. If the information is to remain accessible as long as paper, preservation must be a front-end, rather than an *ex post facto*, action. With paper-based information, publishers, librarians, and archivists can simply warehouse texts, preferably in research repositories and under environmental controls, and then wait to see if research demands develop over the years sufficient to justify the cost of reformatting or other preservation steps. With electronic data, if one simply waits for user communities to emerge, the means to convert data are likely to be long since unavailable by the time that the next cycle of users emerges and new questions for old texts are posed.

Consequently, the preservation of electronic texts is first and foremost a matter of editorial and administrative policy rather than of techniques or materials. Those who create texts, both authors and their publishers, must take a central role in considering long-term accessibility. Among the most important editorial questions to be faced at the establishment of an electronic publication is an existential one: Are the publication and the texts that it contains worthy of long-term accessibility, or is the information only of transient, even if critical, value? As a practical matter, unless the electronic publication has in place a mechanism to ensure ongoing access at least as long as what would be available through printing on acid-neutral paper, the publisher is *de facto* saying that the author's work does not have long-term value even if it merits rapid electronic dissemination. By placing his or her work in an electronic publication without plans for long-term accessibility, the author is saying that the work does not have permanent value.

Not all electronically published products of scholarly research merit long-term accessibility. For example, an electronic bulletin board posting summaries of recently completed research may have little long-term value if its function is to be a prepublication notice of items to appear later in paper or electronic texts or journals of record. What is needed in all cases is a difficult editorial decision on the continuing value of information. Nevertheless, with electronic information, archivists and librarians have to face the difficult collection development or appraisal issue of assessing value without the benefit of time to see the evolution of scholarly writing and research.

Because of the considerable costs involved in creating and maintaining the infrastructure that permits ongoing access to electronic publications, and because of the extent to which commercial and market considerations have so dominated publisher and library relations, action on these items will require a fundamental reconsideration of the relations among authors, publishers, librarians, and archivists. Authors will need to consider future accessibility of their work when they choose a publisher. Publishers' responsibilities will have to go beyond the traditional role of simply distributing texts and instead extend into providing for the disposition of texts. Otherwise they will be delivering products of increasingly dubious utility because of their rapid obsolescence.

Librarians and archivists can play an important role in new relationships with publishers and information providers. Both are aware of the ways people use information especially for purposes well beyond those envisaged as the audience by the author or publisher. In a more structured relationship with publishers and authors, archives and libraries might serve as depositories for either hard copy "dumps" of electronic texts or as predesignated repositories for access systems that will enable long-term research access to electronic texts. With careful up-front planning involving publishers, their boards, librarians, and archivists, the mechanisms to allow long-term access can be established. However, a continuance of the *laissez-faire* system that has evolved for traditional paper and commercial publication will ensure the loss of authors' texts and extraordinary costs for libraries and archives as they strain to meet users' inquiries.

Ultimately, the preservation of electronic publications cannot be solely an archival issue but an administrative one that can be addressed only if the creators and publishers take an active role in providing resources necessary to ensure that ongoing accessibility is part of initial system and product design. Although substantial costs might be involved, there are important administrative incentives to incorporating preservation/continued accessibility into initial production. Furthermore, scholarly publishers and text-based vendors will be able to improve the attractiveness of their product when they can offer assurances of indefinite future access. Certainly some commercially oriented electronic publishers who make no provision for maintaining texts may be able to offer cheaper products, but they will in effect be offering nothing better than the electronic equivalent of newsprint for works that both authors and users believe have an enduring, if not timeless, value.

Electronic publishing has an undisputed advantage in providing rapid and broad distribution, but unless it provides assurances for ongoing accessibility, it will not be able to fulfill key substantive and

“political” roles of scholarly publishing. At its most basic, the purpose of scholarly publishing, whether electronic or paper, is to disseminate an authentic account of research findings for use by the present and for examination and re-interpretation by the future. The very terminology used for key publications as “journals of record” or “archival proceedings” reflects this critical need for scholarly publications to provide a permanent record of the development of knowledge in each discipline. In practical terms, “publications of record” also play a central role in the academic process, especially in the evaluation of scholars for appointment, tenure, and promotion. Even after the issue of peer review for electronic publications is resolved, enduring accessibility of a scholar’s publications will be an important factor if authors and publishers expect electronic publications to be credible elements in promotion reviews. Without publications and journals of record, colleges and universities will be hard pressed to ensure the quality and integrity of their faculties. This is not to suggest that merely ensuring indefinite future access to electronic texts will improve the academic evaluation process, but unless ongoing access is considered, it seems unlikely that electronic publication can be a significant means to disseminate key scholarly texts.

TECHNICAL STEPS

When preservation issues are pushed to the fore during the planning and development of electronic publications and when they are appreciated as matters of administrative and editorial policy, the resolution of technical questions actually becomes simplified. If publishers start with the understanding that they have to provide for long-term accessibility of texts, they will resolve access questions long before current hardware and software become obsolete. Preservation needs can then be met through standard management practices such as daily system backups, off-site storage of backup tapes, thorough system documentation, and quality control over purchases of hardware and storage media. In all those cases where the published text is deemed to be of such ongoing interest that it will be continually provided online, preservation should be indistinguishable from daily system maintenance.

In those cases where the published text may not merit the costs of continuous live and broad access, additional work will be needed. First, through an editorial and management process, involving authors as well as librarians/archivists, the long-term value of the text will have to be assessed, and the best mechanisms to permit continuing accessibility will have to be identified. The archival model of records disposition scheduling can be of enormous utility here because it lays out a mechanism for systematic review and scheduling of the disposition of information.

Technologically, the most critical element for the preservation of, and ongoing accessibility to, electronic texts is the adherence to standards in designing the structure and content of the texts. Since ongoing access will necessitate the movement of text from one system to another to maintain currency of hardware and software, information will have to be encoded in a readily transferable format. The most promising way to achieve this transportability is through use of the Standard Generalized Markup Language (SGML) which was accepted as a standard by the International Standards Organization in 1986. Although seemingly clumsy on first examination, SGML is actually quite flexible and has the enormous advantage of supporting not just the encoding of text but also formatting information and data linkages (e.g., hypertext). Because it is a standard, it allows transportability across hardware and software platforms (Bradley, 1992, pp. 271-74; 1994, p. 10).

An encouraging development is that SGML has been considered to be a critical element for electronic publishing because of its transportability and because it supports multiple representations of a single text (e.g., the Association of American Publishers [1987] has produced a guide for the mark-up of documents for typesetting and printing). This illustrates the convergence of production and archival issues as both publishers and librarians consider the fundamental changes brought about by electronic information systems. In the process, both the nature of preservation and publication are being reconsidered. The principles and mechanisms are already available to resolve these issues to the mutual benefit of authors, publishers, and librarians. Still critically needed are the plans and resources to ensure that the future accessibility of information is secured before products are marketed. This is no small challenge since the momentum of market-driven publishing and feature-laden technology continues to be so overwhelming that the greatest emphasis appears to be on rapid issuance and marketing of products.

ACCESS AND USER SERVICES

User access to archival material may involve a conflict between the rights of privacy and confidentiality and the public rights of freedom of information. Underlying all questions of access is the fundamental consideration of cost. The concept of free access to all published information in the public domain is in opposition to the constitutional protection extended to corporate or individual rights to compensation for the use of intellectual property. Examples of the complexity of such conflicts are afforded by public subsidies of for-profit information services and the private marketing of public information. The marketability and rapid dissemination of electronic publications also provides new and fertile ground for litigants.

CONCLUSION

The publication of electronic information for network access represents a contemporary information delivery system. Archivists must be aware of the process and consider authenticity in the appraisal of information. They must also preserve the content of contemporary information systems and provide access for current and future users. In these activities, they will have to confront and employ rapidly changing technologies, face legal issues surrounding authenticity and property rights, recognize the necessity for the early incorporation of preservation measures into information systems, and serve user clienteles that expect rapid access to archival holdings.

REFERENCES

- Association of American Publishers. (1987). *Reference manual on electronic manuscript preparation and markup*, version 2.0. Washington, DC: AAP.
- Bradley, N. (1992). SGML concepts. *ASLIB Proceedings*, 44(7/8), 271-274.
- Brichford, M. J. (1977). *Archives and manuscripts: Appraisal and accessioning*. Chicago, IL: Society of American Archivists.
- Child, M. S. (1993). *Directory of information sources on scientific research related to the preservation of sound recordings, still and moving images and magnetic tape*. Washington, DC: Commission on Preservation and Access.
- Cuddihy, E. F. (1980). Aging of magnetic recording tape. *IEEE Transactions on Magnetics*, 16(4), 558-568.
- Graham, P. S. (1994). *Intellectual preservation: Electronic preservation of the third kind*. Washington, DC: The Commission on Preservation and Access.
- Mitchell, T. W. (1956). The Illinois records management survey: A new concept in records management. *Illinois Libraries*, 38(6), 139-142.
- National Archives. (1978). Intrinsic value in archival materials. *National Archives Staff Information Papers*, 21, 1-6.
- Neil, B. (1994). What is SGML? *Information Technology and Libraries*, 13(10), 10.
- Stielow, F. (1992). Archival theory and the preservation of electronic media: Opportunities and standards below the cutting edge. *American Archivist*, 55, 332-343.