Through the development of the Genome Data Base, OMIM, and other projects, the laboratory has demonstrated that knowledge management represents a "practical working alternative to existing roles and relationships in the creation and management of scholarly knowledge."[162] Lucier will expand his work in the development of the new Center for Knowledge Management at the University of California at San Francisco.

This section reviewed some of the library community's strategies. The next section recommends actions that the archival profession can take to respond to changing research methods. These actions are an important step toward confronting the transformation of scholarly practice that is as imminent as the new millennium.

CONCLUSION AND RECOMMENDATIONS

The scholarly use of information technology is resulting in dramatic changes in research practices. Essentially two trends are evident: one toward end-user computing and the other toward connectivity. To an increasing extent, social scientists and humanists are performing their own computation in the context of ever greater connectivity. The scholarly use of computers and communication technology for research and information exchange has both short-term and long-term ramifications for archival practice. In the short term, the archival profession needs to address the increasing prominence of network-mediated scholarship. In the long term, the role of the archival profession in the development of next-generation archives that operate in

conjunction with global networks needs to be defined. The following recommendations suggest concrete actions the archival profession can take to address both of these issues during the next decade:

- Establish a presence on the Internet/ NREN.
- Make source materials available for research use over the Internet.
- Create documentation strategies to document network-mediated scholarship and the development of research and education networks as a new communications medium.
- Develop archival methods suitable for operation with NREN.
- Take user practices and computational capacity into account in establishing policies on the management of software-dependent records.
- Recognize and reward initiatives that advance (a) the archival management of electronic records; (b) the response to scholarly use of information technology; and (c) a network-mediated archival practice.

These recommendations are considered in the three-part discussion below.

Part I: Establishing a Network-Mediated Archival Practice

The archival profession, first and foremost, must respond to the emergence of network-mediated scholarship. New methods of searching for sources, communicating with colleagues, disseminating research findings, and providing instruction suggest that scholarly communication is increasingly mediated through electronic networks. The existing Internet and the future NREN represent the new meeting ground where scholars turn for bibliographic information, scholarly dialogues and feedback, the most current publications in their fields, and high-level educational offerings. Increasingly, full-text versions of journals, magazines, newsletters, and even

---

[162]"Knowledge Management: A Collaboration of Academic Scholars, Scientists and Librarians," unpublished statement on the three-year project sponsored by the Council on Library Resources, The William H. Welch Medical Library, Laboratory for Applied Research in Academic Information (15 July 1990).

primary sources are available through networks. In response to this new phenomenon, the archival profession needs to establish a presence on research and education networks and to evaluate the implications of new forms of scholarly communication for standard archival practice.

But before attempting to introduce policy or collaborative action, the archival profession must start using the networks. Indeed, the use of networks is the chief action archivists can take in response to changing patterns of scholarly communication. A presence on the Internet is essential if archivists are to establish credibility as legitimate network collaborators.

Establishing an archival presence on the networks is affordable. Telecommunications hookups involve a modem, communication software, and an e-mail address provided through a link to an already existing network connection. For archivists who do not already possess a modem and who choose not to use public domain communication software, the cost entails a one-time expenditure of, at most, several hundred dollars. Ongoing connect charges in the United States are minimal. Most archivists should experience little trouble obtaining an electronic mail address because the majority of campuses are already wired for network connections, as are federal and state agencies and many private organizations and corporations, especially those affiliated with scientific research and development. In fact, several hundred archivists[163] already participate on BITNET in the network list *Archives and Archivists*. Once hardware and communication software are in place, the archival profession can become an Internet participant.

Recommendation 1: Archivists should begin monitoring and responding to scholars' intellectual activities conducted on networks.

Besides the standard methods for keeping current on research trends, archivists should participate in scholarly electronic conferences. To participate, one signs up, or "subscribes," to a conference. Because thousands of conferences exist, archivists should use conference lists and compiled directories to select those that involve subject areas most closely approximating the holdings of their repository. For instance, a repository strong in women's history sources may subscribe to the lists devoted to women's and gender studies. An institution noted for its collection of pre-Civil War holdings may choose a conference devoted to eighteenth century America. Social welfare archives may sign up for conferences related to social work, social activism, and family studies. Those with strong collections of utopian records may select the Shaker conference. Repositories noted for their holdings on the arts may join the many conferences on theater, film, and drama.[164]

One way scholars use these conferences is to exchange information about source materials related to research topics. In an effort to participate in these dialogues, NARA's Center for Electronic Records began monitoring several scholarly conferences in 1991. The conferences offer the center a forum for responding to several dozen additional inquiries each month from scholars and librarians relating to the center's holdings. One center staff member currently spends about thirty minutes each day monitoring four BITNET Listservs on topics related to government documents,

---

[163]As of June 1992, approximately 440 archivists subscribed to the Bitnet Archives and Archivists listserv.

[164]Examples of electronic conferences are from Kovacs, *Directory of Scholarly Electronic Conferences,* 3rd rev.

electronic data sets, social science data lists, and the Vietnam War.[165]

These conferences not only provide a means for keeping up with trends in scholarly research but also provide a mechanism for establishing a presence on the networks by attaching a name and institutional affiliation to each communication. As simplistic as this sounds, a more substantive involvement with networks can occur only when archivists are familiar with the arena's discourse and techniques and when the archival profession is established as a network participant. We therefore recommend as an initial action that archivists establish a presence on the Internet by participating in network conferences.

Recommendation 2: The archival profession should identify and implement archival methods appropriate to new forms of scholarly communication.

Establishing a presence on the networks is a necessary first step. But in addition to conference participation, the archival profession should pursue archival methods responsive to changes in scholarly communication. These new archival practices and techniques include: providing access on the Internet to source materials in machine-readable form, initially as bit-mapped images; documenting the activities of network-mediated scholarship; and establishing archives that operate in the Internet/NREN environment.

2 (a): The archival profession should make source materials available on the Internet. The archival profession should make sources directly available to scholars via research and education networks. The sources should include both records that originate in electronic form and those created in nonelectronic forms. Since the transfer of nonelectronic records to machine-readable form is a formidable undertaking, this discussion focuses primarily on conversion strategies.

Converting nonelectronic sources to machine-readable form is justified for several reasons. First, the scholarly expectation that full-text materials should be available online as a research convenience is unmistakably evident and growing.[166] Indeed, electronic document delivery represents the undisputed standard for the information field. Second, beyond convenience, conversion of source materials to machine-readable form is essential for analyses that rely on computational processing. Third, with increasing frequency, the types of questions posed by researchers require entire electronic libraries of sources, instead of a single collection, available for computational processing. From this perspective, the larger the corpus of converted collections, the greater the research value.

As further justification, in the absence of an archival role in the conversion of source materials, the commercial sector is certain to prevail. This is not to suggest that many types of conversion projects would not be more suitable as commercial sector undertakings. But as the transition to the online era proceeds, archivists have the responsibility to ensure that publicly available records remain so when converted to machine-readable form and to alert citizens to the danger of losing the right of free access through inaction.

The proposal to convert source materials to machine-readable form is neither radical nor original. Many leaders in the library profession argue that conversion is one of

---

[165]Conversation between Avra Michelson and Ted Hull of NARA's Center for Electronic Records, 26 August 1991; also Ted Hull, "NNXA Reference Report," NARA Center for Electronic Records, June 1991, draft.

[166]Shrinking travel allocations also may spur requests for online access, if the cost of geographically dispersed archival research exceeds academic budgets.

the most important actions librarians can take to establish a comprehensive record of scholarship.[167] As discussed earlier, some libraries are already performing pilot conversions. Further, the Commission on Preservation and Access recently released several reports recommending that preservation microfilming include the generation of digital images.[168] In an alternative approach, Cornell University Library, in conjunction with Xerox Corporation and the Commission on Preservation and Access, demonstrated the feasibility of directly converting text to digital form, avoiding the costs associated with microfilming.[169]

In arguing that archivists should convert nonelectronic holdings to machine-readable form, we are not suggesting that it is either feasible or desirable to convert all records. The volume of archival holdings is simply too great, and many holdings do not warrant the investment. Rather, our point is that it is time to begin breaking the tie with the printed past and establishing a connection with the machine-readable future.

Converting source materials to machine-readable form entails the resolution of many issues that are beyond the scope of this paper. However, we would like to comment on a few basic archival questions related to conversion: What should be converted? What electronic form should conversion result in? What kind of new descriptive devices are necessary to facilitate the independent use of electronic versions of source materials?

What should be converted? Most repositories periodically, if not regularly, microfilm deteriorating collections of enduring value. Applying current technology, microfilm preservation projects could be expanded or transformed to digital conversion projects through the development of several funded, model programs. The benefit of establishing digital conversion programs based on preservation microfilming is that many procedures in place for microfilming are also suitable for imaging. First, materials for preservation microfilming typically are selected because they are in need of preservation attention and are deserving of wider access. These two elements are adequate criteria for the current selection of collections to be digitized.[170]

---

[167]See, for instance, Smith, *The Librarian, the Scholar, and the Future of the Research Library,* 71–72. Clifford Lynch also recommends conversion of source materials to digital form in "Achieving the Promise: A Proposed Strategic Agenda for Libraries and Networked Information Resources in the 1990s," unpublished paper presented at the Networks for Networkers II Pre-Conference, Chantilly, Virginia 17–19 December 1990, 18 (Also published under that title in *Networks for Networkers: Critical Issues for Libraries in the National Network Environment,* edited by Barbara Evans-Markuson with Elaine W. Woods [New York: Neal-Schuman Publishers, forthcoming]).

[168]See Donald J. Waters, *From Microfilm to Digital Imagery* (Washington, D.C.: Commission on Preservation and Access, June 1991), and Michael Lesk, *Image Formats for Preservation and Access* (Washington: D.C.: Commission on Preservation and Access, July 1990). These reports explore microfilming as a means to achieve digitization.

[169]The Cornell project, co-managed by Anne R. Kenney and Lynne K. Personius, involves the direct conversion of one thousand volumes of brittle books to digital form. Half of the volumes are mathematical books, some of which are handwritten or contain formulas and graphic images. The Cornell project uses Xerox hardware that is capable of producing both digital output and enhanced print output from a digital copy. This collaborative effort has produced meaningful data on costs, procedures, and models associated with digitization programs useful to the archival profession. See Kenney and Personius, "The Future of Digital Preservation."

[170]Other categories of records also may make good candidates for conversion even though they are not deteriorating. In selecting records primarily to provide greater access, other factors should be considered, including the nature and extent of use of the records, the institutional visibility or impact afforded by the conversion, the type of image required for use, the volume and condition of the records requiring conversion, special labor costs, and the extent to which conversion can be accomplished through scanning, optical character recognition (OCR), or manual input. But we think it would be a mistake for the archival profession to expend much effort at this point on refining selection criteria until the results of a number of digital conversion projects can be analyzed. Fur-

Second, the document preparation processes used with microfilming are largely compatible with digital conversion.[171] This means that handling procedures in place for preservation microfilming can essentially be applied to digitization. Third, microfilming and digitizing can be intertwined technical processes. That is, while it is technically possible to generate a microfilm copy as output from a digitized collection, it is also possible to generate a digitized copy of a record set from microfilm output. This means that it is possible to create microfilm and then digitize the output, or digitize directly and then generate microfilm. As such, repositories concerned with the longevity of digital storage mediums, or their ability to move digital data from one generation of technology to another, can continue to rely on microfilm for preservation purposes and still convert records to machine-readable form.

Repositories that plan to microfilm are encouraged to establish pilot digital programs that draw on many structures already in place for preservation microfilming. The archival profession needs tested models to establish the most cost-effective procedures for administering ongoing conversion programs. Pilot projects should provide sufficient technical and programmatic guidance and an awareness of how digital sources are used, to equip the profession with the ability to implement large-scale digital conversions.

What electronic form should digital conversions result in? The profession's assessment of appropriate electronic forms will probably change over time. The overriding concern, however, must be to identify the kinds of representations patrons need. Do they need a facsimile image of documents? A stream of straight ASCII text that can be manipulated? ASCII text encoded with tags that identify document structures and formats? Although the electronic forms that patrons need depends on the type of research they are conducting, very little is known about the actual use of electronic documents for different types of research.

Trends in the technology suggest that in the future the archival profession should be able to provide access to electronic sources both as bit-mapped images and encoded text. But current limitations make large-scale encoding of text an unrealistic undertaking. For many reasons, the existing methods of performing ASCII conversions, manual key entry or automatic optical character recognition (OCR) are inadequate. For example, the cost of performing key entry with great volumes of materials is prohibitive, and OCR processes are unreliable with handwritten script and unusual type fonts. In contrast, bit-mapped conversions, which result in image representations (like facsimiles, but potentially of far greater resolution), are readily attainable with today's technology. Further, automatically converting bit-mapped images of modern printed documents to ASCII is typically considered a straightforward process (equivalent to OCR). If desired, encoded text can be generated from the ASCII version, provided the relevant structural information has been retained.

ASCII and encoded text differ from bit-mapped images in that the latter cannot be searched and computationally processed without considerable programming. It is highly probable, however, that software designed to encode text automatically will improve and reduce in cost during this decade. If this happens, it may be feasible to justify large-scale textual encoding. Until

---

ther, the Commission on Preservation and Access has contracted with Margaret Child to reconsider current criteria used to select source materials for preservation microfilming. Presumably the archival profession will find the results of this study relevant to digital conversion efforts as well.

[171]See Archival Research and Evaluation Staff, *Optical Digital Image Storage System: Project Report* (Washington, D.C.: National Archives and Records Administration, 1991), 6; and Kenney and Personius, "The Future of Digital Preservation," 9.

then, sources not amenable to OCR should be converted to bit-mapped images. But since bit-mapped images will not satisfy the research needs of certain scholars, archivists should monitor advances in OCR and structure-encoding software.

What kind of new descriptive devices are necessary to facilitate independent use by researchers of electronic versions of source materials? Digital versions of large archival collections will need specialized finding aids, descriptors, navigational aids, or informational hooks to facilitate their independent use.[172] Developing these finding aids and navigational tools represents a key challenge for the information profession. Nonetheless, it would be ill-advised to convert unstructured and voluminous collections to machine-readable form, or to make collections that originate in machine-readable form available for independent use, without addressing the need for a descriptive system suitable to the electronic environment. As a further complication, standard bibliographic approaches to retrieval are proving an inadequate method for locating and managing remote electronic text banks. But metadata, data about data that archivists typically collect about a body of records, may serve as the basis for a supplementary descriptive system to complement existing bibliographic information. Administrative histories, accession records, and other contextual data used to establish the provenance of a collection may prove very useful in retrieving information from electronic sources in the absence of human intermediaries.[173]

It is encouraging to note that contextual information accreted to each document in records originating in machine-readable form is likely to be greater than in their print counterparts. For example, e-mail messages interchanged on the Internet identify the sender and institution, the receiver(s) and institution(s), the date and time of transmittal, and the subject of the communication. Archival intervention into the design phase of software could result in the accumulation of other metadata that would be useful for both accountability and retrieval purposes. We therefore endorse the National Historical Publications and Records Commission's proposal to research the implications of capturing and retaining data, descriptive information, and contextual information in electronic form, and we speculate that the findings of this research can also advance the development of descriptive systems suitable for independent use by end-users.[174]

2 (b): Archivists should develop and implement a strategy for documenting network-mediated scholarship as a new phenomenon of scholarly communication. A key finding of this report is the substantial level of scholarly activity being conducted outside the purview of traditional archival practice. Network-mediated scholarship raises two very different but related documentation issues for the archival profession. The first is the need to document the origin and administration of research and education networks themselves. The second is the need to document the

---

[172]For a justification of the need for new access tools, see Clifford A. Lynch and Cecilia M. Preston, "Internet Access to Information Resources," in *Annual Review of Information Science and Technology (ARIST)*, vol. 25, edited by Martha E. Williams (Amsterdam: Elsevier Science Publishers B.V., 1990); and Lynch, "Achieving the Promise," 24–25.

[173]Charles Robb, at the Kentucky Department for Libraries and Archives, is developing a locator system for statewide information using metadata to comple-

ment existing bibliographic information. This approach may be useful in developing descriptive systems that provide access to information within a collection as well. See Charles Robb, "Networking Metadata in Kentucky," unpublished paper presented to the National Association of Government Archivists and Records Administrators, Chicago, July 1990.

[174]This recommendation is part of *Research Issues in Electronic Records*, 10–11. Charles Dollar also argues that archivists should define metadata elements in his report, *The Impact of Information Technologies on Archival Principles and Methods*, 98–100.

programmatic use of these networks for the advancement of scholarship and learning. As an approach and process, documentation strategy[175] represents a tool that archivists can use to address these documentation problems—e.g., to identify the key agents operating in the network environment, to determine the universe of documentation that exists, and to develop recommendations for preserving documentation of enduring value.

The large number of agents and the global scope of activities associated with research and education networks suggests that archivists may want to collaborate and seek multi-institutional funding for documentation projects. At least three types of archival repositories are well-positioned to initiate such projects: (1) college and university archives, because network research and education efforts originate largely in academia; (2) government archives, because government is a key partner in most academic-based collaborative research projects and network-mediated education programs (either as a funder, research associate, or network administrator); and (3) discipline history centers (such as the American Institute for Physics, the Beckman Center for the History of Chemistry, and the Babbage Center), as these centers, by definition, explore a universe of documentation and are heavily devoted to science and technology, disciplines in which network-mediated scholarship is currently the most pervasive.

The documentation effort should identify key representatives to participate in strategic discussions, such as those from the Internet and scholarly communities, academic computing centers, private industry, and government research laboratories. A goal of the effort should be to clarify the principal records-creating agents and the activities that warrant preservation. The project report should include a statement on the nature of electronic archival records and the relationship of these sources to non-electronic documentation.

This recommendation involves a certain urgency because existing documentation tends to be transient. In fact, compilers of several network directories report that at least a half dozen recent scholarly electronic conferences are already defunct, as are more than a dozen electronic newsletters and journals.[176] Some argue that these efforts become inactive when moderators switch jobs and no longer possess the equipment or time to continue in that role or when the interest in a once-timely topic, such as the Gulf War, dissipates. Instead of papers removed to an attic for storage, the records of a defunct electronic conference typically take the form of a mass of bits abandoned on a campus mainframe computer or file server, awaiting a purge of the file by a systems administrator in a routine cleanup. Given this situation, academic computing staff represent key contacts for campus archivists concerned with network files. State archivists also should be concerned with the transient nature of network communication because network-mediated distance education programs are under way in most state departments of education.[177] In summary, archivists at insti-

---

[175]Two seminal essays that together provide an intellectual foundation for the concept of documentation strategy, as well as an examination of procedures and case examples are: Helen Willa Samuels, "Who Controls the Past," *American Archivist* 49 (Spring 1986): 109–24; and Larry Hackman and Joan Warnow-Blewett, "The Documentation Strategy Process: A Model and a Case Study," *American Archivist* 50 (Winter 1987): 12–47.

[176]Correspondence via Bitnet on 23 August 1991 between Avra Michelson and Diane Kovacs, compiler of *Directory of Scholarly Electronic Conferences;* also, a list of defunct electronic journals and newsletters appears in Michael Strangelove, *Directory of Electronic Journals and Newsletters.*

[177]See two reports by Barbara Kurshan: *Statewide Telecommunication Networks: An Overview of the Current State and the Growth Potential* (Roanoke, Va.: Educorp Consultants, December 1990), and with Marcia Harrington, *Statewide Education Networks: Survey Results* (Roanoke, Va.: Educorp Consultants, April 1991). Both are available through Bitnet from the author (Kurshan@vtvm1.bitnet).

tutions that support online scholarly communication are urged to seek funding for programs to identify and preserve valuable records related to the administration of networks comprising the Internet and network-mediated scholarship.

2 (c): The archival profession should support the development of archives designed to operate on global networks. The growth in network-mediated scholarship suggests that the archival profession needs to define its role in relation to the development of archives designed to operate in the global network environment. The need for archival operations on research and education networks is already widely recognized by the network community. For example, program planning in the network community involves archival concerns. At a biannual meeting of the Coalition for Networked Information (CNI), many subcommittees reported on work that entailed the resolution of archival functions in a network environment.[178] Although separate from the archival profession, CNI represents a group that is identifying issues related to the archiving of network resources.

Further, the development of electronic network archives is already evident. Most moderators of scholarly electronic conferences maintain an archives of the conference's transactions accessible via the network.[179] Others are capturing subject-oriented transactions across research and education networks and making the archives available on the Internet.[180] Still

others are exploring commercial models for preserving both volume and breadth in network transactions.[181] Those involved in network archiving communicate with one another through electronic conferences about such issues as data compression algorithms, information filtering techniques, and file transfer protocols.[182] This means that seminal models for microarchiving within a network environment are already in place, while those for archiving on a grander scale are either on the drawing board or being prototyped, each established apart from the work of the traditional archival profession.

Archivists must not underestimate the significance of these actions. The future of the archival mission in relation to electronic communication is being defined by a set of agents wholly separate from the work of the traditional archival profession. Further, the scope of the new archival agents is apt to grow as NREN evolves into a piece of the backbone used in the conduct of official government business.[183] The appropriate role for the archival profession in this arena remains undefined, but the key questions are clear. Can the archival profession establish the political authority necessary to improve the archival methods used in conjunction with research and education network transactions, and can it rise to the

---

[178]Observation by Avra Michelson at the CNI spring meeting, 18–20 March 1991, Washington, D.C.

[179]Correspondence by Avra Michelson with Diane Kovacs, compiler of the *Directory of Scholarly Electronic Conferences,* on 23 August 1991 via BITNET.

[180]For instance, Edward Vielmetti at MSEN in Ann Arbor, Michigan, collects and makes available descriptions of network resources publicized on the networks [emv@msen.com]; Nathan Torkington at the Computing Services Center in Wellington, New Zea-

land, maintains a publicly accessible electronic archives of text on information management captured from network exchanges [gnat@kauri.vuw.ac.nz].

[181]Vielmetti has developed commercial models for archiving select network transactions.

[182]The key electronic conference where these issues are discussed is comp.archives.admin moderated by Edward Vielmetti.

[183]The U.S. Office of Personnel Management recently released guidelines already effective for the acceptance of electronic signatures in the conduct of official government business. With the issue of electronic signatures resolved, the use of networks in official government business can be expected to increase rapidly. See U.S. Office of Personnel Management, *Federal Personnel Manual,* Chapter 293, Subchapter 6, Installment 39, 1 April 1991 (Washington, D.C.: U.S. Government Printing Office, 1991).

challenge of defining an archival practice suitable not only for electronic records but also for a new communication medium?

Part II: Establishing a Strategy for the Future Usability of Electronic Records

No discussion of information technology trends can ignore the issues surrounding the storage and use of electronic records themselves. Although this subject has been discussed in the archival literature,[184] our focus here is on the scholarly research perspective. This article has concentrated on the near-term effects of information technology on current scholarly practice and products. It is equally important, however, to consider how new ways of producing records (whether they are of scholarly origin or not) will affect future users of those records. In particular, how will the creation of electronic records affect future scholars when they use such records in their research? What current technology trends bear on the ways these future scholars will perform their research and—by implication— on the ways future archives will have to serve them?

One of the main advantages of electronic

information is that it is usually digital, which ensures that it can be copied and transmitted without loss or degradation. Yet, ironically, the preferred media on which this digital information is stored—disk, tape, and even CD-ROM—have far shorter shelf lives than acid-free paper or microfilm. Moreover, these media tend to become unusable long before they reach their ultimate age limits. As technology evolves, it quickly reaches a point where older media can no longer be accessed by existing equipment. It is only somewhat facetious to express this irony by saying that digital data lasts forever—or five years, whichever comes first. There is no theoretical problem with storing digital information on archival media, including microfilm, but such media are not in popular use, nor does evidence suggest that they will become so. This problem has a straightforward, though cumbersome and relatively expensive, solution: to "update" or "migrate" data, that is, to copy the data from one medium to another as media wear out or become obsolete. Although various technology trends (including the continued development of optical storage devices such as CD-ROM) may improve the longevity of media, the overall trend of continued improvement and replacement of media implies that the problem of obsolescence is unlikely to disappear in the foreseeable future.

Despite this problem, it is axiomatic that the records produced by governments, organizations, individuals, and researchers themselves will become increasingly "electronic" over the next few decades. This implies that scholars of the not-so-distant future will be confronted increasingly with electronic records as both the primary and secondary source materials for their research. Moreover, the current first generation of such records will have unique historical significance, representing the most drastic change in the form and conception of records since the introduction of print-

[184]A selection of the key literature includes David Bearman, *Archival Methods,* Archives and Museum Informatics Technical Report, no. 21 (Spring 1989); Advisory Committee for the Co-ordination of Information Systems (ACCIS) *Management of Electronic Records: Issues and Guidelines* (New York: United Nations, 1990); U.S. House, Committee on Government Operations, "Taking a Byte out of History: The Archival Preservation of Federal Computer Records," House Report 101-978 (Washington, D.C.: Government Printing Office, November 1990); *Research Issues in Electronic Records* (St. Paul, Minn.: published for the National Historical Publications and Records Commission, Washington, D.C., by the Minnesota Historical Society, 1991); David Bearman, ed., *Archival Management of Electronic Records,* Archives and Museum Informatics Technical Report no. 13 (Pittsburgh: Archives and Museum Informatics, 1991); Margaret Hedstrom, "Understanding Electronic Incunabula: A Framework for Research on Electronic Records," *American Archivist* 54 (Summer 1991): 334–54.

ing, or even of writing.[185] Yet at the current rate of technological change, electronic documents (and the programs that produce and access them) typically become obsolete and unusable in a distressingly short time. How can the loss of this unique generation of records be prevented? How will scholars be able to understand and analyze these documents decades from now? How can archives hope to preserve such documents in a form scholars will be able to use?

Furthermore, media longevity is only a part—and in many ways the easier part—of the problem. Migrating data can keep them "accessible," but to be usable they must be more than just accessible: they must also be interpretable. The data stored on digital media are simply binary digits (bits), which cannot be interpreted without a translation of the codes they represent and an understanding of the structure in which they are placed on their media. Migrating data may solve the media longevity problem, but by itself it does not solve the larger problem. Like an illiterate monk dutifully copying text in a lost language, migration may save the bits but lose their meaning. Even if we assume that the media longevity problem can be solved, what technology trends bear on whether electronic records will be interpretable in the future?

This issue is often referred to as that of *software-dependent records*, though there is somewhat more to the problem than this term suggests. Software-dependent records are electronic documents that can be read only by using some particular piece of computer software (that is, a program). Examples of software-dependent records include documents created with word processing or electronic publishing programs, spreadsheets, databases, geographic information systems (GISs), and

hypertext/hypermedia. Though a data file for such a document may be saved on some medium (such as a disk), the file can be properly interpreted only by its software; the document itself is accessible (and in some cases may come into existence) only by running the software.[186] This can be thought of as the problem of "preserving" electronic documents. However, in this case, "preservation" means more than simply preserving media; unlike printed records, electronic records require software and hardware in order to be accessed and interpreted.

The obvious way to access a software-dependent document is to run the software that produced it. However, programs themselves quickly become obsolete, and running obsolete software is currently very difficult. Any given program works only on certain computers and only with certain system software. This means that accessing a document may actually require the user to run this entire hardware and software environment. In fact, what is typically meant when a document is called "software-dependent" is that it can be accessed only by running the entire hardware and software environment in which it was created. The problem is that such environments become obsolete in the blink of an archival eye, and maintaining them in working condition beyond that time is a complex, costly, and ultimately futile task.[187] Preserving elec-

---

[185]See Jay David Bolter, "Text and Technology: Reading and Writing in the Electronic Age," *Library Resources and Technical Services* 31 (January/March 1987): 12–23.

[186]In a very real sense, *all* electronic documents are software-dependent. Simple text and numeric files are not typically referred to as "software-dependent" only because they are encoded and stored in fairly straightforward ways that currently are considered obvious (e.g., simple sequences of ASCII codes representing characters). Yet even these cannot be accessed or interpreted without hardware and software that can understand their encoding.

[187]For several discussions of this issue, see David Bearman, *Collecting Software: A New Challenge for Archives and Museums,* Archives and Museum Informatics Technical Report no. 2 (Pittsburgh: Archives and Museum Informatics, 1987, reprinted 1990); and Coalition for Networked Information Director Paul Evan

tronic documents in a way that will allow future access to their form and meaning is therefore not straightforward.

There appear to be two general approaches to providing meaningful future access to software-dependent documents. Either they must be transformed in some way that makes them independent of the software that created them, or they must be saved along with some kind of description of their associated software sufficient to allow accessing them as was originally intended. The first approach might be facilitated by the development of standards for various kinds of documents, whereas the second approach might be facilitated by the development of formal models of computation. Several technology trends bear on each of these approaches.

Software-dependent documents might be preserved in a usable form by transforming them so that they become "software-independent" in some way. For each recognized category of program now in use (word processing, database, spreadsheet, etc.) a standard data file format might be defined, along with a standard set of functions that any such program can perform. For example, most word processing programs provide functions for displaying pages of text, footnotes, and chapter headings. In principle, a data file for a document from any such program could be transformed into some standard format, and its behavior could be duplicated by some standard program.[188] This transformation process would

have to be repeated periodically as the standard itself evolved. Standardization trends such as those discussed above may help make this possible. However, there may always be programs whose behavior cannot be duplicated by any standard or which do not even fit into the recognized categories of programs (e.g., word processing or database). As noted above, standards generally lag behind the advancing technology; until computer science becomes far better formalized (that is, based on firm, theoretical underpinnings), there will always be programs that defy the most well-conceived efforts at standardization. Policies in various organizations may attempt to force the use of programs that conform to standards, but current trends of technological innovation make enforcement difficult because users find it hard to resist new capabilities, whether they are standard or not.

Even aside from standardization efforts, a "natural migration" of documents occurs as the programs on which they depend evolve through successive versions. New versions of programs often provide upward compatibility to allow old documents to migrate into the required updated forms. It may be possible, as has been suggested,[189] to rely to some extent on this phenomenon to keep documents accessible. The effectiveness of this approach, however, is limited by the fact that periodic upheavals occur in software paradigms. Two examples of

---

Peters, "The Machine Aspects of Preservation," unpublished paper (ca. 1990).

[188]SGML is an attempt to provide a standard for this kind of text, though it is generally recognized that even a standard for text will not magically remove all the incompatibilities among existing word processing formats. Another example of this approach that has been discussed in the literature involves relational databases. The argument has been made that a database produced by *any* relational database management system (RDBMS) can be transformed into a standard form that can be used by any other RDBMS. See the National Archives and Records Administration's re-

sponse to the recommendations in "Taking a Byte Out of History," and Kenneth Thibodeau, "To Be Or Not to Be: Archives for Electronic Records," in *Archival Management of Electronic Records,* edited by David Bearman, 1–13. Although this may be true to a large extent, it is a relatively atypical example; relational database systems are one of the very few higher level applications for which a formal (mathematical) computational model exists. Most other common applications, such as word processing, spreadsheets, hypertext/hypermedia, or GISs are not nearly this well formalized.

[189]Dollar, *The Impact of Information Technologies on Archival Principles and Methods,* Chapters 1-4, draft version.

such upheavals are the change from simple textual tables to spreadsheets and the change from hierarchical databases to relational databases. Such upheavals make it difficult enough to transform documents that are crucial to the daily functioning of organizations; transforming old documents that are no longer in use may require more effort than most organizations are willing to spare.

The alternative to transforming software-dependent documents into software-independent form is to interpret them by somehow using the software that they depend on, despite its being obsolete. Interpretation does not necessarily require actually running the software. If a complete description existed of how a program interprets its data files in accessing a document, it would not be necessary to save the software itself (or its environment). The document could be accessed by following this description, effectively recreating the *behavior* of the software. In most cases, unfortunately, such complete descriptions of software exist *only* in the form of the software itself. Computer science is not yet very good at describing what complex software does.[190]

Interpreting a software-dependent document by using the software it depends on therefore requires either being able to run the software that has been saved along with the document (by effectively recreating its environment), or interpreting the software without running it (effectively recreating, or emulating, its behavior). The former option requires saving vast (though finite)

documentation for the software and its environment, including detailed technical descriptions of any required hardware and all of its components.[191] The latter option requires a more sophisticated computational theory than is currently available, i.e., an understanding of the semantics of what programs do at the human level of information processing and how they do it. Without such a theory, it remains impractical to interpret software except by running it in its original hardware and software environment.[192] Current trends toward improving the formal specification of systems and environments may facilitate the former option, whereas trends toward modeling human level computational processes may facilitate the latter.[193] Finally, it should be noted that the overriding trend toward increased computational power may enable the performance of tasks that now appear unthinkable, just as we now routinely perform computations that were unthinkable a decade or two ago. Such future tasks might include automatically decoding lost file structures, transforming obsolete document formats through successive generations of standards, or recreating the behavior of ar-

---

[190]There are exceptions to this, such as the relational database case discussed above. In general, however, current formal descriptive techniques cannot capture the "human level" semantic behavior of programs. What is required is a computational theory, not of how programs *work*, but of what they *do* for their users; i.e., a theory of human information processing that describes such things as how humans create and use documents and how humans interact with each other to perform research.

[191]Although this is a huge task, it may not be insurmountable: These environments could not exist in the first place if they did not already possess such technical descriptions. Furthermore, many of these descriptions are already in patent or copyright offices, where they might be accessible for this purpose.

[192]Recreating the behavior of a program by figuring out what it was intended to do and building a new program that does what the original program did is sometimes called "reverse engineering." It is widely recognized as a difficult task.

[193]Advances in computational theory may enable future generations of scholars to understand how we viewed and manipulated our documents far better than we understand it ourselves. The present is, after all, only the dawn of the information age, and the organizing principles of the new "computation" paradigm are only beginning to emerge. Future scholars may have a far better formal (i.e., mathematical) understanding of computation and human information processing; this would provide them with a theoretical framework that could explain any kind of software-dependence and allow them to reconstruct past capabilities at will.

chaic computational environments from imperfect documentation. These computational possibilities may well allow future generations of scholars to derive the equivalent standard form of obsolete software-dependent documents in their archives or to reproduce the behavior of the software that produced them at will.

In the context of scholarly research and information technology, the issue of software-dependent records can be phrased in terms of two questions: "How can access to software-dependent documents be provided to future scholars?" and "What technology can help to provide this access?"

To answer the first question, one must articulate certain assumptions about what kinds of access future scholars are likely to need to such documents and what they will do with them after they have accessed them. The software used to create a software-dependent document determines the capabilities available to its author for viewing and manipulating it. How accurately must scholars be able to reproduce these capabilities? Is it enough to preserve the content of such a document without its form? Is it enough to preserve its content and form without being able to recreate the way its author saw it?[194] These questions require making assumptions about the kinds of research future scholars will perform, which can be informed by analyzing trends in scholarly practice, as undertaken above. Given such assumptions, how would alter-

native software-dependent records management policies constrain or enhance the capabilities of future scholars in performing their research using software-dependent documents?[195]

To answer the second question, one must articulate other assumptions about the technological future (while recognizing that all such assumptions are speculative).[196] In particular, what do current technology trends imply about future capabilities for accessing software-dependent records?

Saving data files for software-dependent documents is a necessary but insufficient step toward making them usable. As discussed above, data can be migrated to new media to keep them readable, but data must be more than just readable to be usable: They must also be interpretable. Is there some way to transform such documents before saving them in archives, so that they can be used *without* their software? If so, what would this sacrifice in terms of being able to recreate the author's original capabilities? Alternatively, is there some practical way of saving the software with each document (in particular, without maintaining obsolete hardware/software environments) so that the software itself can be used in the future to access the document? If solutions to these problems are not found and implemented soon, much of the first generation of electronic documents—representing a unique historical event in the evolution of records—will be irretrievably lost.

To summarize, there appear to be two general approaches to solving this problem, as discussed in the archival literature: Transform each document and save it in software-independent form, or save the software for each document in some way

---

[194]Margaret Hedstrom suggests that "The solution to preservation of electronic records lies somewhere between the present approach of preserving only data values and the need to retain all of the functionality of an active records system. There are tremendous advantages to retaining the descriptive, search, retrieval, and manipulation functions of some automated systems. The ability to retain more complex electronic records and more of the useful functionality of automated systems, however, will remain beyond the control of archivists if they continue to utilize only the tactics [that] have been employed in the past." "Archives: To Be or Not to Be: A Commentary," in *Archival Management of Electronic Records,* 28.

[195]This article raises this question without attempting to answer it. Our point is that the assumptions that underlie any answer must be made explicit.

[196]The archival literature on this subject has not yet generally articulated such assumptions.

that allows it to be used in the future to interpret the document.

The solutions that have been proposed in the literature for both approaches (e.g., translating documents into one of a few current standard forms or keeping hardware/software environments running for as long as possible) appear to be based on implicitly conservative assumptions about future technology. It seems likely, however, that inevitable advances in computational theory and computational power will produce a vastly more capable future, enabling better, longer-range solutions to one or both of these approaches. This analysis has implications for the actions that should be taken now to ensure the preservation of these records. We see the following recommendation as a necessary step toward deciding on such actions.

Recommendation 3: The archival profession should establish an evolving policy on the management of software-dependent records, informed by an assessment of the kinds of access future scholars will require to such records and a realistic assessment of the computational capabilities that will be available in the future.

Because of the short effective life of most electronic media and the rapidity with which software-dependent documents tend to become obsolete and unusable, this recommendation has an urgent aspect: Electronic records of enduring value that are not appropriately preserved will soon be lost to posterity.

The archival profession should take steps to ensure that its evolving software-dependent records management policy considers the ways that future scholars are likely to use these records and the ways that future technology is likely to facilitate this use. Assessments, such as the one we have undertaken here, which attempt to analyze trends in scholarly practice and information

technology should be used to attempt to project future needs and capabilities that are *realistic*, i.e., neither wishfully grandiose nor unimaginatively chained to the past. These projections should be used to produce evolving policies aimed at the moving target that is the future.

Evolving trends in scholarly practice should be sought out by the archival profession, in an attempt to coordinate the development of archival policies with the perceptions and projections of those scholars who represent the leading edge of change in scholarly research practice. This coordination might be achieved through scheduling paper sessions or panel discussions on evolving scholarly practice, to be presented at archives and library science conferences and at conferences in various scholarly disciplines. Workshops, journals, or network discussions might also be organized on this subject, soliciting input from scholars while establishing the archival profession as a focal point for this inquiry.

Similarly, archivists should seek out evolving trends in technology, with particular emphasis on formalisms and standards for representing various kinds of documents and on formal models of computation and human information processing, which ultimately may make it possible to describe the behavior of software in ways that will allow it to be emulated in the future. In this endeavor, archivists should actively engage the computer science community as a partner, for example by organizing sessions or panels on these subjects at both computer science and archives conferences.

Finally, archivists should engage in an ongoing effort to understand the most likely future uses of software-dependent records, and they should articulate their assumptions about future scholarly practice and future computational capability as a prerequisite for proposing archival policies on the management of software-dependent records.

Part III: Recognizing and Rewarding Leadership

Recommendation 4: The archival profession should reward activities that advance archival practice with information technology, electronic records, and electronic communication.

The archival profession must respond to the changing patterns of scholarly communication and the emergence of a new communication medium. Leadership capable of guiding the archival profession should be cultivated by promoting graduate education programs, collaborative projects, and professional coalitions targeted at advancing archival operations in global network environments. The Society of American Archivists and the field's other professional associations should recognize and reward excellence in research, pilot projects, collaborative associations, and programmatic implementations related to the management of electronic records, the use of information technology to improve archival practice, and the establishment of archival methods suitable to modern communication mediums.

BIBLIOGRAPHY

Adman, Peter. "Computers and History." In *Information Technology in the Humanities: Tools, Techniques and Applications.* edited by Sebastian Rahtz, 92–103. New York: Halsted Press, 1987.

Advisory Committee for the Coordination of Information Systems. *Management of Electronic Records: Issues and Guidelines.* New York: United Nations, 1990.

Alberico, Ralph. "The Development of an 'Information Superhighway'." *Computers in Libraries* 10 (January 1990): 33–35.

Alexander, Michael. "But Can You Read It Like a Book?" *Computerworld* 24 (19 November 1990): 18.

Anderson, Paul A. "Using Artificial Intelligence to Understand Decision Making in Foreign Affairs: The Problem of Finding an Appropriate Technology." In *Artificial Intelligence and National Security,* edited by Stephen J. Cimbala, 133–46. Lexington, Mass.: Lexington Books, 1987.

Archival Research and Evaluation Staff. *Optical Digital Image Storage System: Project Report.* Washington, D.C.: National Archives and Records Administration, 1991.

Arms, Caroline, ed. *Campus Strategies for Libraries and Electronic Information.* Bedford, Mass.: Digital Press, 1990.

———. "Libraries and Electronic Information: The Technological Context, Part Two." *EDUCOM Review* 24 (Fall 1989): 34–43.

Arms, William Y., and Thomas J. Michalak. "Carnegie Mellon University." In *Campus Strategies for Libraries and Electronic Information,* edited by Caroline Arms, 243–73. Bedford, Mass.: Digital Equipment Corporation, 1990.

Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing. *Proposal for Funding for an Initiative to Formulate Guidelines for the Encoding and Interchange of Machine-Readable Text.* Prepared for the National Endowment for the Humanities, 1988.

Baerson, Kevin M. "Marines Put Library Online." *Federal Computer Week* 5 (2 September 1991): 1, 4.

Bailey, Charles W., Jr. "Electronic (Online) Publishing in Action . . . *The Public-Access Computer Systems Review* and Other Electronic Serials." *Online* 15 (January 1991): 28–35.

———. "Intelligent Multimedia Computer Systems: Emerging Information Resources in the Network Environment." *Library Hi Tech* 8 (1990): 29–41.

Bantz, David A. "The Values of the Hu-