Digital Preservation: A Time Bomb for Digital Libraries

Author(s): Margaret Hedstrom

Source: *Computers and the Humanities*, 1997/1998, Vol. 31, No. 3 (1997/1998), pp. 189–202

Published by: Springer

Stable URL: http://www.jstor.com/stable/30200423

Digital Preservation: A Time Bomb for Digital Libraries

MARGARET HEDSTROM*

School of Information, University of Michigan, Ann Arbor, Michigan 48109-1092, U.S.A.

Abstract. The difficulty and expense of preserving digital information is a potential impediment to digital library development. Preservation of traditional materials became more successful and systematic after libraries and archives integrated preservation into overall planning and resource allocation. Digital preservation is largely experimental and replete with the risks associated with untested methods. Digital preservation strategies are shaped by the needs and constraints of repositories with little consideration for the requirements of current and future users of digital scholarly resources. This article discusses the present state of digital preservation, articulates requirements of both users and custodians, and suggests research needs in storage media, migration, conversion, and overall management strategies. Additional research in these areas would help developers of digital libraries and other institutions with preservation responsibilities to integrate long-term preservation into program planning, administration, system architectures, and resource allocation.

1. The Challenges of Digital Preservation

The purpose of preservation is to protect information of enduring value for access by present and future generations (Conway, 1990: 206). Libraries and archives have served as the central institutional focus for preservation, and both types of institutions include preservation as one of their core functions. In recent decades, many major libraries and archives established formal preservation programs for traditional materials in paper, microform, photographic, and to a lesser degree audio-visual formats. Preservation programs include administrative and technical

---

* Margaret Hedstrom is an Associate Professor at the School of Information, University of Michigan where she teaches in the areas of archives, electronic records management, and digital preservation. Before joining the faculty at Michigan in 1995, she worked for ten years at the New York State Archives and Records Administration where she was Chief of State Records Advisory Services and Director of the Center for Electronic Records. Dr. Hedstrom holds a Ph.D. in History from the University of Wisconsin in Madison. She has published widely on many aspects of archival management and electronic records and she is currently writing a book on digital preservation. Her current research interests include digital preservation strategies, the impact of electronic communications on organizational memory and documentation, and remote access to archival materials. She is a fellow of the Society of American Archivists and was the first recipient of the annual Award for Excellence in New York State Government Information Services.

components, such as hiring staff with expertise in preservation administration, using preventive measures to arrest deterioration of materials, taking remedial actions to restore the usability of selected materials, and incorporating preservation needs and requirements into overall program planning and resource allocation. Preservationists within the library and archival community have been instrumental in developing an array of tools and methodologies to reduce the decay of traditional materials and to restore books and documents that have deteriorated to such an extent that their longevity and usability are threatened. Provisions for fire protection and adequate environmental controls frequently are incorporated into new library and archival facilities. Rehousing of acid-based paper materials is a common task in many repositories, and microfilming is used extensively as a cost-effective means to preserve endangered materials. Undertakings such as the Brittle Books Initiative, the American Newspapers Project, and the National Endowment for the Humanities (NEH)-funded microfilming program have saved millions of unique and imperiled items (*Preserving the Intellectual Heritage*). Many libraries and archives have curbed their voracious appetites for acquiring new materials in an effort to balance the breadth and depth of their holdings against long-term stewardship responsibilities. Pressure from the preservation community provided the catalyst for many publishers to change over from acidic to acid-neutral paper in the production of published works. Introducing more stable materials at the beginning of the information production process represents in a significant victory for preservation interests which in the long run will reduce the need for salvage efforts.

Much remains to be done to preserve cultural, intellectual, and scholarly resources in traditional formats that form the foundation for humanities research and teaching. An estimated 80 million embrittled books reside in American libraries, 10 million of which are unique; and countless journals, newspapers, photographs, and documents require preservation treatment to survive well into the next century. Thousands of repositories lack the means for disaster prevention or adequate environmental controls to avoid catastrophic loss of their holdings. The success stories and regular use of established preservation methods are found almost exclusively in developed countries. Within developed countries, major institutions tend to have the most extensive preservation programs and preservation techniques are better developed for print materials than for special formats such as photographs, film, and video (*Preservation of Archival Materials*).

Digital preservation adds a new set of challenges for libraries and archives to the existing task of preserving a legacy of materials in traditional formats. I define digital preservation as the planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable. I intentionally use the term "continuing" rather than "permanent" value to avoid both the absolutism and the idealism that the term "permanent" implies (O'Toole). My concept of digital preservation encompasses material that begins its life in digital form as well as material that is

converted from traditional to digital formats. Although preservationists have been battling acid-based papers, thermo-fax, nitrate film, and other fragile media for decades, the threat posed by magnetic and optical media is qualitatively different. These new recording media are vulnerable to deterioration and catastrophic loss, and even under ideal conditions they are short lived relative to traditional storage media. They are the first reusable media and they can deteriorate rapidly, making the time frame for decisions and actions to prevent loss a matter of years, not decades.

More insidious and challenging than media deterioration is the problem of obsolescence in retrieval and playback technologies. Innovation in the computer hardware, storage, and software industries continues at a rapid pace, usually yielding greater storage and processing capacities at lower cost. Devices, processes, and software for recording and storing information are being replaced with new products and methods on a regular three- to five-year cycle, driven primarily by market forces. Digital works which are created using new or emerging software applications are especially vulnerable to software obsolescence because standards for encoding, representation, retrieval, and other functions take time to develop. New formats also tend to be more complex because they can handle a wider variety of information representations and information processing functions. New multimedia applications have to handle multiple representations of text, data, images, sound, and motion as well as manage the relationships among the components of complex digital objects. Records created in digital form in the first instance and those converted retrospectively from paper or microfilm to digital form are equally vulnerable to technological obsolescence.

Digital preservation is constrained by the absence of established standards, protocols, and proven methods for preserving digital information and by the tendency to consider preservation issues only at the end of a project or after a sensational loss. With few exceptions, digital library research has focused on architectures and systems for information organization, retrieval, presentation, and visualization, and on the administration of intellectual property rights (Levy and Marshall). The critical role of digital libraries and archives in ensuring the future accessibility of information with enduring value has taken a back seat to enhancing access to current and actively used materials. As a consequence, digital preservation remains largely experimental and replete with the risks associated with untested methods. Digital preservation requirements have not been articulated from either the user or provider perspective, nor have they been factored into the architecture, resource allocation, or planning for digital libraries.

The juxtaposition of mass storage and long-term preservation offers fertile ground for discussing conceptual and methodological challenges facing digital libraries and archives. The two terms "mass storage" and "long-term preservation" embody a contradiction in the current state of affairs of digital library development, representing a time bomb that threatens the long-term viability of this new type of knowledge resource. New technologies for mass storage of digital information

abound, yet the technologies and methods for long-term preservation of the vast and growing store of digital information lag far behind. The strategies, methods, and technologies for long-term preservation of digital information that do exist are not necessarily feasible technologically for preservation on a mass scale, nor are they affordable given the vast quantities of digital information being generated. Our ability to create, amass, and store digital materials far exceeds our current capacity to preserve even that small amount with continuing value.

2. Digital Preservation Requirements

In order to preserve digital materials on a scale commensurate with mass storage capabilities and in formats that are accessible and usable, it is necessary to articulate some basic requirements. There are two different perspectives on digital preservation requirements: those of users of digital materials and those of libraries, archives, and other custodians who assume responsibility for their maintenance, preservation, and distribution. Libraries and archives will not accomplish their preservation missions if they do not satisfy the requirements of their users by preserving materials in formats which enable the types of analyses that users wish to perform. At the same time, libraries and archives may not be able to satisfy all requirements of all potential users due to resource constraints, competing priorities, and lack of technical expertise. By making preservation requirements explicit from both the users' and custodians' perspectives, libraries and archives will be better able to balance competing demands and to integrate digital preservation into overall planning and resource allocation.

Potential information needs are varied, unpredictable, and almost endless (Gould). Any generalizations, even if restricted to one community of users such as humanities scholars, run the risk of overlooking and understating potential user needs. Although precise requirements vary among disciplines, some basic needs transcend fields and disciplines. The ability to establish the authenticity and integrity of a source is critical to all users, regardless of whether the source was created by an individual, generated in the conduct of institutional business, or produced through a formal publication process (Lynch). Mechanisms that will enable users to establish the origin, provenance, and authenticity of digital documents require archives and libraries to preserve contextual and descriptive information in addition to the content of digital documents. Attributes such as formal document structures, metadata that document the maintenance and use history of the document, time and date stamps, and references among documents are essential for determining authenticity, understanding the provenance of sources, and placing them in a larger context (Graham). While these requirements are important as well for traditional materials, they are even more significant for digital materials which are easily altered, copied, and removed from their original context.

Michelson and Rothenberg (1992) argue that networking and access to digital sources will change all dimensions of the scholarly work process, including

identifying sources, communicating with colleagues, interpreting and analyzing data, disseminating research findings, and teaching. If their projections are correct, digital preservation programs must enable a high degree of integration between source material and analytical processes by coupling research sources with the tools necessary to analyze them; by maintaining linkages between research results and the sources on which they are based; and by providing a means to incorporate primary sources into teaching. Users will seek documents that are easily retrieved and manipulated, transmittable, and transportable from a repository to the sites of research, presentation, and teaching. It seems safe to assume that humanities scholars will need the capability to search through and select relevant sources from large bodies of heterogeneous materials, to compare sources with each other, and to view specific documents at high levels of granularity. Digital preservation will add little value to the research process if it serves only as an alternative form of storage from which analog replicas are produced for use with conventional analytical methods. Preserving digital materials in formats that are reliable and usable, however, will require long-term maintenance of structural characteristics, descriptive metadata, and display, computational, and analytical capabilities that are very demanding of both mass storage and software for retrieval and interpretation.

Archives, libraries, and other types of repositories that are struggling to meet escalating user expectations with limited financial and technical resources may well express digital preservation requirements differently from end users. From the perspective of a repository, storage systems should be capable of handling digital information in a wide variety of formats, including text, data, graphics, video, and sound. Digital storage is not simply an alternative means for storing print formats. Many types of digital objects do not have print equivalents and cannot be preserved in non-digital formats. Ideally, storage media will have a long life expectancy, a high degree of disaster resistance, sufficient durability to withstand regular use, and very large storage capacities. Conversion from analog to digital formats and migration to new generations of technology will be rapid, accurate, and inexpensive enough to permit very large scale transfers of heterogeneous materials. Storage space requirements will be minimal and not demand highly sensitive environmental controls. To make digital preservation affordable to the widest possible range of organizations and individuals, equipment, media, and maintenance costs must be modest.

3. Current Preservation Strategies and Their Limitations

Most librarians and archivists have accepted the basic wisdom that digital preservation depends upon copying, not on the survival of the physical media (Lesk). But copying, also referred to as "refreshing" or "migration," is more complex than simply transferring a stream of bits from old to new media or from one generation of systems to the next. Complex and expensive transformations of digital objects often are necessary to preserve digital materials so that they remain legit-

imate representations of the original versions and useful sources for analysis and research (*Preserving Digital Information*). Current methods for preserving digital materials do not fully support achieving these objectives. When faced with the responsibility for preserving digital materials, archives and libraries face a series of complex and difficult choices based on the format of the original materials, the anticipated uses for it, and the technical and financial resources available to invest in preservation initiatives which range from very elementary and established methods to proposals that have not yet been tested.

Transferring digital information from less stable magnetic and optical media by printing page images on paper or microfilm is probably the most commonly used preservation strategy. It seems ironic that just as libraries and archives are discovering digital conversion as a cost-effective preservation method for certain deteriorating materials, much information that begins its life in electronic form is printed on paper or microfilm for safe, secure long-term storage. Yet, high-quality acid neutral paper can last a century or longer while archival quality microfilm is projected to last 500 years or more. Paper and microfilm have the additional advantage of requiring no special hardware or software for retrieval or viewing. Perhaps this explains why in many digital conversion projects, the digital images serve as a complement to rather than a replacement for the original hard copy materials (Conway, 1994).

Another strategy for digital preservation is to preserve digital information in the simplest possible digital formats in order to minimize the requirements for specific retrieval software and to avoid problems of software obsolescence. Digital information can be transferred across successive generations of technology in a "software-independent" format as ASCII text files or as flat files with simple, uniform structures. Several data archives hold large collections of numerical data that were captured on punch cards in the 1950s or 1960s, migrated to two or three different magnetic tape formats, and now reside on optical media. As new media and storage formats are introduced, the data can be migrated without any significant change in their logical structure. This approach has the distinct advantage of being universal and easy to implement. It is a cost-effective strategy for preserving digital information in those cases where retaining the content is paramount, but display, indexing, and computational characteristics are not critical. As long as the preservation community lacks more robust and cost-effective migration strategies, printing to paper or film and preserving flat files will remain the methods of last resort for many institutions and for certain formats of digital information.

Libraries and archives with large, complex, and diverse collections of digital materials are only beginning to test strategies that normalize various types of holdings by converting digital records from the great multiplicity of formats into a smaller, more manageable number of standard formats ("Pilot Accessioning Projects Summary Report"). A repository might accept textual documents only in one or a few commonly available commercial word processing formats or require that documents conform to standards like SGML (ISO 8879). Stable and standard-

ized methods for mark-up, such as the use of SGML as a tool for representing the logical structure of texts, can preserve structural, presentation, and retrieval capabilities without becoming dependent on the longevity of particular software applications. Databases might be stored in one or a few common formats or converted to a SQL (Structured Query Language) compliant format, while image files might conform to the tagged image file format (TIFF) or another common image format with standard compression algorithms.

This approach has the advantage of preserving more of the display, dissemination, and computational characteristics of the original materials, while reducing the large variety of customized transformations that otherwise would be necessary to migrate material to future generations of technology. The strategy rests on the assumption that software products which are either compliant with widely adopted standards or are widely dispersed in the marketplace are less volatile than the software market as a whole. Most common commercial products today provide utilities for backward compatibility and for swapping documents, databases, and more complex objects between software systems. Although this strategy simplifies migration and may lower digital preservation costs by reducing the amount of customized reformatting needed as technology changes, it does not eliminate the need for regular migration of digital materials. Both software and standards continue to evolve and even repositories with structurally homogeneous holdings can expect to be required to migrate their digital materials periodically.

Current methods fall far short of what is required to preserve digital materials. All current preservation methods involve trade-offs between what is desirable from the standpoint of functionality, dependability, and cost and what is possible and affordable with current technologies and methods. Consequently, most repositories are coping by employing interim and less than desirable strategies, if they are addressing digital preservation issues at all. For example, the simplicity and universality of printing to paper or microfilm come at the expense of great losses in the retrieval and reuse potential of digital information. Migration strategies that involve reformatting of digital materials to a simple standard format usually eliminate the structure of documents and relationships imbedded in databases. Computation capabilities, graphic display, indexing, and other features often are lost, thus limiting future analytical potential. Normalization to standard formats is not always technically feasible and it usually is quite costly. Archives and libraries must also contend with entirely new forms of electronically-enabled discourse and new forms of artistic and cultural expression that do not have predecessors in the analog world. No current preservation method is adequate for preserving dynamic data objects from complex systems. There are no established conceptual models or technical processes for preserving multi-media works, interactive hyper-media, on-line dialogues, or many of the new electronic forms being created today. The archival requirements to preserve content, context, and structure of digital documents and the need to maintain the capability to display, link, and manipulate

digital objects to satisfy user requirements only heighten software dependency in digital repositories.

The preservation community is only beginning to explore possible alternatives to storing digital information in "software-independent" form. Rothenberg (1995) proposed an approach for maintaining the content of digital materials intact without losing the ability to retrieve meaning-rich sources. He recommended retaining the original document in its original format encapsulated in a virtual "envelope" that contains software instructions for retrieval, display, and processing of the message in the envelope. The envelopes would contain contextual information and the transformational history of each object. Execution of the instructions would rely on an archive of hardware and software emulators or on instructions in the envelope with specifications on how to construct emulators.

The digital preservation strategies discussed so far are based on the assumption that libraries and archives will be passive recipients of digital materials. A complementary strategy seeks to place preservation concerns farther upstream in the process of generating digital information by developing or recommending standards for the format, structure, and description of digital objects so that they are more amenable to long-term preservation. Research on electronic records management requirements has defined metadata standards for evidence that will support the need for integrity, authenticity, reliability, and archiving through standards for "metadata encapsulated objects" (Bearman and Sochats; *Bulletin of ASIS*). Archives, libraries, and other institutions with preservation responsibilities will benefit if the information systems which generate digital information are designed to support long-term preservation needs, such as migration, adequate description, and linkages between the content of digital materials and their larger organizational or intellectual context. Wide scale adoption of data and communication standards by the originators of digital information to support their current business needs may also facilitate long-term preservation. Rapid implementation of electronic commerce depends on widespread development and adoption of standards for electronic data interchange and secure transactions. Many organizations are adopting standards for formats and definitions to enable exchange, reuse, and sale of digital information and to reduce conversion and maintenance costs. Standards initiatives that address business needs for the secure and reliable exchange of digital information among the current generation of systems will impose standardization and normalization of data that ultimately will facilitate migrations to new generations of technology. Yet to benefit fully from the synergy between business needs and preservation requirements, cultural heritage concerns should be linked to equally critical social goals, such as monitoring global environmental change, locating nuclear waste sites, and establishing property rights all of which also depend on long-term access to reliable, electronic evidence.

Much research and development is needed before libraries and archives can ensure that even a small portion of our digital heritage will survive. It is fair to say that the state of development in digital preservation remains largely experimental.

Only a few libraries, archives, and other institutions have established digital preservation programs, while most research and innovation comes from pilot projects and prototypes. Tested methods that have been proven effective on a small scale in a limited number of repositories are not feasible for preservation of many of the types of digital materials that archives and libraries will confront in their preservation endeavors.

4. Areas for Research and Development

The current state of digital preservation suggests several fruitful areas for research and development. I will discuss four areas: storage media, migration, conversion, and management tools. These four domains are mutually dependent and ultimately must to be integrated into an infrastructure for digital preservation. Yet better solutions are necessary in all four areas before such integration can occur. Finally, I will share some observations about the issues of scale and cost that must be considered if libraries and archives are going to achieve any degree of systematic preservation.

5. Storage Media

The limited life of magnetic and optical media pose a significant problem, although this is not the primary limiting factor for digital preservation. Recent research on the longevity of magnetic media indicate a useful life span of 10 to 30 years if they are handled and stored properly. Some optical disk technologies promise life spans of up to 100 years. Most authorities argue that enhanced media longevity is of little value because current media outlast the software and devices needed to retrieve recorded information. As Conway has stated, the chain of interrelated elements required for digital preservation is only as strong as its weakest link (Conway, 1996: 13). Nevertheless, improvements in the stability, capacity, and longevity of the base storage media are needed to drastically reduce the vulnerability of digital materials to loss and alteration and to decrease storage and maintenance costs. Ample research and experience provide evidence of what can go wrong with magnetic media as a result of binder degradation, magnetic particle instabilities, and substrate deformation (Van Bogart). Optical media are susceptible to damage from high humidity, rapid and extreme temperature fluctuations, and contamination from airborne particulate matter (U.S. National Archives and Records Administration). To prevent these problems, it is imperative to store magnetic and optical media under strict environmental controls that are not always available, affordable, or convenient. Even modest improvements which produce storage media with larger per unit storage capacities and greater tolerance to variations in temperature and humidity will lower preservation costs by lessening the need for strict environmental controls, reducing the frequency with which digital media must be "refreshed" through recopying, and decreasing the number of storage units that must be handled.

This raises the question, however, of whether research on incremental improvements in current storage technologies will benefit preservation in the long run or whether alternative approaches to digital storage will meet archival requirements more adequately. As a frame of reference it is worth remembering that microfilm, which is considered the only acceptable archival storage medium, lasts at least 500 years with minimal maintenance if stored properly. New storage technologies, such as the experimental High-Density Read-Only Memory (HD-ROM) technology, which uses an ion beam to inscribe information on pins of stainless steel or iridium, may be worth investigating. The HD-ROM is capable of storing 180 times more information than current CD-ROM technology at roughly one-half of one percent of CD-ROM costs. According to one release about this technology, the HD-ROM is impervious to material degradation and it requires no bit stream interpreter because the technology can describe in human-readable form all of the instructions needed to interpret the data (LANL Ion Beam Storage). Such an approach illustrates the potential for solutions built on entirely new storage technologies.

6. Migration

Better methods for migration of digital materials to new generations of hardware and software are much needed for digital preservation regardless of breakthroughs in mass storage technologies. Planning for migration is difficult because there is limited experience with the types of migrations needed to maintain access to complex digital objects over extended periods of time. When a custodian assumes responsibility for preserving a digital object it may be difficult to predict when migration will be necessary, how much reformatting will be needed, and how much migration will cost. There are no reliable or comprehensive data on costs associated with migrations, either for specific technologies and formats or for particular collections, and little research underway on methodologies that would reduce the costs and burdens of migration.

Organizations with preservation responsibilities would benefit tremendously from the development of backward compatibility paths that would be included as a standard feature of all software. Backward compatibility or migration paths would enable a new generation of software to "read" data from older systems without substantial reformatting and without loss of retrieval, display and computational capabilities. Although backward compatibility is increasingly common within a particular software product line, migration paths are not commonly provided between competing software products or for products that fail in the marketplace.

Stewards of digital material have a range of options for preserving digital information. One might preserve an exact replica of a digital record with complete display, retrieval, and computational functionality, or a representation of the record with only partial computation capabilities, or a surrogate for the record such as an abstract, summary, or aggregation. Detail or background noise might be dropped

out intentionally through successive generations of migration, and custodians might change the format or storage media. Enhancements are technologically possible through clean-up, mark-up, and linkage, or by adding indexing and other features. These technological possibilities in turn impose serious new responsibilities to present digital materials to users in a way that allows them to determine the authenticity of the information and its relationship to the original source. Methods to document changes in digital objects during their life span need to be incorporated as an integral part of improved migration methods.

There are few well developed methods for preserving and migrating software so that it might be used to recreate digital documents that have the "look and feel" of the original sources. Maintaining repositories of obsolete hardware and software has been discussed periodically, but usually dismissed out of hand as too expensive and not demonstrably feasible. This approach deserves more serious consideration as a strategy for maintaining continuing access to certain types of digital materials. Feasibility studies and cost/benefit analyses should be conducted to determine the technological, economic, and commercial feasibility of maintaining selected legacy software systems and performing specialized migrations or, alternatively, of building and maintaining software emulators. Such an approach would support replay of original sources and contribute to the preservation of software as a significant cultural and intellectual resource in its own right.

7. Conversion

Faster, cheaper, and higher resolution conversion technologies are another critical element needed to make digital preservation feasible on a large scale. Most archivists and librarians accept the fact that we live in a hybrid environment where paper, microfilm, video, and magnetic and optical media need to interoperate in a more integrated and transparent manner. The vast majority of primary sources today still reside on paper and/or microfilm with little chance that we will see the mass conversion of existing archival and library holdings to digital formats. Research and planning for digital preservation must recognize that repositories will be dealing with conversion for a long time and that investments in improving capture rates, accuracy, resolution, and verification will have long-term benefits. Moreover, improvements in conversion technologies may support hybrid solutions to preservation and access problems by permitting repositories to store certain formats of digital material on stable media, such as microfilm, with on demand conversion to digital form for analysis and reuse. Efforts to capture and store descriptive mark-up on film for subsequent conversion are hampered by unacceptable error rates in OCR technology and cumbersome conversion processes (Giguere).

8. Management Tools

A fourth area for research is the development of management tools for digital libraries and archives that integrate descriptive control and maintenance with digital

storage systems. Dynamic digital objects, such as those found in hypertext systems, pose special management problems for both current and future retrieval and reuse. The boundaries of hypertext sources, like those found on the World Wide Web today, are difficult to ascertain because no single party or institution controls changes in all of the documents and links that make hypertext objects live and highly responsive information resources. A high degree of volatility accompanies these objects because the contents of documents change, the sites where information resources are stored change, and the links between documents change, move, and vanish. Some recent tools, such as "crawler" software which is capable of traversing a portion of the Web and noting maintenance problems such as broken links, moved documents, modified documents, and objects that have exceeded their expiration dates, have potential applications in managing large digital archives (Ackerman and Fielding; Kahle). If these tools were augmented to address preservation problems, they have the potential to serve as filters for appraisal and selection, to monitor repositories for physical deterioration, and to identify objects that are threatened because of technology obsolescence.

Research and development of tools that would imbed more intelligence about the preservation status of digital material into the objects themselves would make monitoring and maintenance of large digital collections more automatic. Current methods for monitoring the physical status of digital materials are labor intensive, unreliable, and potentially damaging to the materials themselves. Recommended procedures for monitoring physical deterioration of magnetic media, for example, involve reading a small sample of items periodically to determine whether any data losses have occurred (Eaton). The potential exists to build monitoring and reporting mechanisms into digital objects, storage systems, and network architectures that could support self-reporting of physical status and initiate automatic maintenance procedures.

Despite important differences, some lessons from traditional preservation are transferable to the digital environment. In order to avoid commitments that far exceed available resources and costly rescue and restoration efforts, preservation must become an integral part of the planning, design, and resource allocation for digital libraries and archives. Integration of preservation requirements and methods with access and maintenance systems is essential to manage fully and efficiently the processes of migration, regeneration, and documentation of the life of digital objects. Planning for preservation must become an integral part of the design and management of digital libraries and archives. If left as an afterthought, there is little reason to believe that long-term preservation of digital information will be any more affordable than preservation of conventional formats has been.

In developing new tools and methods for digital preservation, researchers and developers must bear in mind the issues of scalability, affordability, and ease of implementation. The preservation community has at its disposal a variety of tactics for digital preservation which appear to work effectively for specific types of materials in certain restricted environments, but current methods are not scaleable

to the general problem of digital preservation. This is not to suggest that there is or should be a single solution to digital preservation. The methods used will vary depending upon the complexity of the original digital objects, the extent to which the functionality for computation, display, indexing, and authentication must be maintained, and the requirements of current or anticipated users. But any solution must be scaleable from the laboratory, prototype, or pilot project to the wide range of individuals and institutions who have responsibility for ensuring that digital materials last longer than the current generation of technology permits.

Another closely related issue is the question of affordability. Regardless of how the responsibility for digital preservation is distributed, societies only allocate a small and finite amount of resources to preserving scholarly and cultural resources. In the digital environment it seems likely that preservation responsibilities will be distributed among individual creators, rights holders, distributors, small institutions, and established repositories. Decisions made in the production and dissemination process about formats, data structures, standards, storage media, and the like will influence which digital information survives and how much it will cost to maintain it. Therefore, it seems imperative that digital preservation technologies become affordable and accessible to the wide range of individuals and institutions who have some part in the process of preserving digital materials.

Finally, it would be beneficial to both the preservation community and to those conducting research on issues of longevity, migration, and conversion if there were more venues for exchange of ideas, requirements, and recent developments. Without a continuing dialogue between humanists, preservationists, and the scientific community it is difficult to include preservation requirements in scientific research endeavors, and it is challenging for those outside the scientific community to keep up with and evaluate new products. Today, digital preservation strategies are shaped primarily by the requirements and constraints of established repositories which are seeking affordable and practical solutions and methods. In the future, user requirements for more robust and flexible tools for using and analyzing preserved digital resources must also be incorporated into research and development of digital preservation strategies and methods.

References

Ackerman, Mark S. and Roy T. Fielding. "Collection Maintenance in the Digital Library." <http://csdl.tamu.edu/DL95> (1995).

Bearman, David and Ken Sochats. "Metadata Requirements for Evidence." <www.lis.pitt.edu/~nhprc/meta96.html> (1995).

Bulletin of the American Society for Information Science. "Special Issue on Electronic Systems and Records Management in the Information Age." 23(5) (June/July 1997).

Conway, Paul. *Preservation in the Ditigal World.* Washington, D.C.: Commission on Preservation and Access, 1996.

Conway, Paul. "Digitizing Preservation." *Library Journal,* (February 1, 1994), 42–45.

Conway, Paul. "Archival Preservation in a Nationwide Context." *American Archivist,* 53(2) (1990), 204–222.

Eaton, Fynnette L. "The National Archives and Electronic Records for Preservation." In *Preservation of Electronic Formats: Electronic Formats for Preservation*. Ed. Janice Mohlhenrich. Ft. Atkinson, WI: Highsmith Press, 1993, pp. 41–61.

Giguere, Mark D. "Electronic Document Description Standards: A Technical Feasibility of Their Use in the Augmentation of the Microform Preservation of Contextual Cues Embedded in Structured Electronic Documents During Successive Digital/Analog/Digital Reformatting." Ph.D. dissertation, School of Information Science and Policy Studies, State University of New York at Albany, 1995.

Gould, Constance. *Information Needs in the Humanities: An Assessment*. Stanford, CA: The Research Libraries Group, 1988.

Graham, Peter S. "Requirements for the Digital Research Library." *College and University Research Libraries*, 56(4) (July 1995), 331–339.

Kahle, Brewster. "Preserving the Internet." *Scientific American*, 276 (March 1997), 82–83.

"LANL Ion Beam Storage Holds 180 Times More Info than CD-ROMS." *Science and Engineering News*, June 23, 1995, downloaded from HPCwire and redistributed to <PACS-HUPVM1.EDU>.

Lesk, Michael. *Preservation of New Technology: A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access*. Washington, D.C., Commission on Preservation and Access, 1992.

Levy, David M. and Catherine C. Marshall. "Going Digital: A Look at Assumptions Underlying Digital Libraries." *Communications of the ACM*, 58(4) (1995), 77–84.

Lynch, Clifford. "The Integrity of Digital Information: Mechanics and Definitional Issues." *Journal of the American Society for Information Science*, 45(10) (1994), 737–744.

Michelson, Avra and Jeff Rothenberg. "Scholarly Communications and Information Technology: Exploring the Impact of Changes in the Research Process on Archives." *American Archivist*, 55(2) (1992), 236–315.

O'Toole, James M. "On the Idea of Permanence." *American Archivist*, 52(1) (1989), 10–25.

"Pilot Accessioning Projects Summary Report." University of the State of New York, State Education Department, State Archives and Records Administration. Building Partnerships for Electronic Recordkeeping: The Final Report and Working Papers of the Building Partnerships Project. Albany, NY, 1995. <http://unix6.nysed.gov/pubs/build.htm>.

The Preservation of Archival Materials. Washington, D.C.: Commission on Preservation and Access, 1993.

Preserving Digital Information: Report of the Task Force on Archiving Digtal Information. Report commissioned by the Commission on Preservation and Access and The Research Libraries Group, Washington, D.C.: Commission on Preservation and Access, May 1, 1996.

Preserving The Intellectual Heritage: A Report of The Bellagio Conference. Washington, D.C.: The Commission on Preservation and Access, 1993.

Rothenberg, Jeff. "Ensuring the Longevity of Digital Documents." *Scientific American*, 272(1) (1995), 24–29.

U.S. National Archives and Records Administration. Digital Imaging and Optical Digital Disk Storage Systems: Long-Term Access Strategies for Federal Agencies. Technical Information Paper No. 12. National Technical Information Service, Washington, D.C., 1994. <gopher://gopher.nara.gov:70/11/managers/archival/papers/postscri>.

Van Bogart, John W.C. *Magnetic Tape Storage and Handling: A Guide for Libraries and Archives*, Washington, D.C.: Commission on Preservation and Access and the National Media Laboratory, 1995.