



## A Moral and Legal Obligation: Preservation in the Digital Age

DEANNA MARCUM\*

### ABSTRACT

The Commission on Preservation and Access and the Research Libraries Group (both US organizations) created a task force to study the requirements for preserving digital information. The report, published in May 1996, recommends a certification process for digital archives and calls for a legal authority to rescue digital files in danger of being lost. The paper reviews the reactions of the library, archival, and educational communities and describes the current actions being taken.

©1997 Academic Press Limited

Jarislav Pelikan, in *The Idea of the University: A Re-examination*, describes two primary functions for universities. He suggests that support for teaching, research and publication constitutes the “endowment of living genius,” while efforts to preserve or archive knowledge by organizations like libraries, museums and archives represent the “embalming of dead genius.” Librarians, archivists, and museum curators have always accepted their responsibility for seeing that the dead geniuses — those who have created knowledge and left it to the world as a legacy — are available to subsequent generations through their words, visual images, and other creations.

Although preservation has always been difficult to accomplish, largely because financial resources have not matched the speed with which paper deteriorates, preservation is a particularly acute problem today. At least in some of the scholarly disciplines, the conventional published record is not the essence of the intellectual progress of that field. Instead, on-line databases, on-line exchanges of preprints, listservs, and the like constitute the core of scholarly communication. Conventional publication in these disciplines is a redundant process, undertaken to synthesize the state of the

\*President, Commission on Preservation and Access, USA.

art or, lacking other means, to generate a certified archival record of the work.

In such disciplinary fields, the preservation efforts of librarians and archivists must be focused on preserving the electronic materials, rather than the published record. The embalming of dead genius in the pursuit of knowledge and its long term accessibility to others means that we must find ways to preserve information in the form in which it is generated, i.e. digitally.

It was to this end that the Commission on Preservation and Access, in partnership with the Research Libraries Group formed a Task Force on the Archiving of Digital Information. Co-chaired by Don Waters of Yale University and John Garrett of CyberVillages Corporation, the Task Force issued a report in May 1996 that notes that the preservation problem is no longer a situation in which we place brittle books under a microfilm camera in a well defined process. Instead, we are faced with what the Task Force calls "a grander problem of organizing ourselves over time and as a society to maneuver effectively in a digital landscape." Don Waters, in a speech to the Association of Research Libraries, said: "the effort to meet the cultural and economic imperatives of digital preservation requires us to build, almost from scratch, a system of infrastructure for moving the record of knowledge naturally and confidently into the future. The systematic approach on which we need to agree in order to build this infrastructure, has at least two dimensions: the elements of the system and the manner in which we interact to deploy those elements and construct the system and subsystems for digital archiving."

#### CONCLUSIONS

The Task Force recognized that most of the challenges associated with digital preservation are organizational — not technical. Their key findings were:

- The first line of defense against loss of valuable digital information rests with the creators, providers, and owners of digital information.
- Long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives.

- A sufficient number of trusted organizations must exist that are capable of storing, migrating, and providing access to digital collections.
- A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information.
- Certified digital archives must have the right and duty to exercise an aggressive rescue function as a fail-safe mechanism for preserving valuable digital information that is in jeopardy of destruction, neglect, or abandonment by its current custodian.

#### *Recommendations*

The final report focuses on three essential questions: (1) What does digital preservation entail? (2) How do we organize ourselves to do it? and (3) What steps should we take to move forward?

#### *What does Digital Preservation Entail?*

Preservation in both the paper and digital realms included: preserving the content, or the substance of information; preserving the context required to understand and interpret the information well into the future; and providing the means to access and use the information. In each of these functions preservation of digital information is analogous to preserving traditional media, but presents new sets of problems.

#### *Preserving Content*

The first act of preservation in any environment is identifying or selecting material that merits preservation efforts. The principles and judgements required to predict the future need for information in various disciplines are as complex and varied in their application to digital information as they are to more familiar media. The fluidity and dynamic nature of digital data, however, add some new dimensions. The choice of version to retain is more difficult since many more versions of a resource are likely to exist. Interactive and dynamic databases, which change from moment to moment, can only be “preserved” through samples or snapshots. No commonly accepted body of practice yet governs such choices and little experience exists demonstrating the future usefulness for different sampling techniques. Hyper-linked files add further layers of decision-making, as the extent to which linked resources should also be selected and included for preservation must be determined.

The fluidity of digital information and the ease with which it can be changed introduce the need for authentication of the version that is selected for preservation. A repository of digital information must protect information from tampering and must employ techniques for ensuring users that the information held is actually what it is represented as being. A variety of techniques for protecting and authenticating data exist, but their employment is not yet widespread and easily put into production.

In the short term, any provider of digital information must store the data and provide the systems management and engineering to maintain and deliver it, such as back-up, maintenance of redundant files, etc. Longer term retention of files requires additional techniques. Because digital files are dependent on software and hardware to use them, these too must be kept active. Maintaining a museum of unsupported hardware and software platforms is not practical. Data must be migrated to work on new platforms. If data are stored in flat files, that is files in which the data content is easily separable from the software to use it, then this migration is not necessarily difficult. However, increasingly, content and functionality are inseparable, and both must be preserved. Migration then becomes an increasingly complex systems engineering task. One potentially important technique for migration is emulation, that is, development of software that can emulate the environment in which the original software operated. New systems designs and technologies are needed to facilitate this work. To the extent that major software vendors are encouraged to maintain backward compatibility as they develop new versions of their products, the task of migration will be greatly facilitated.

The challenges presented by migration illustrate the fundamentally different conceptual bases that distinguish preservation as it applies to digital rather than analog media. Archival preservation via migration requires a commitment to unknown future activities with unpredictable future costs. This essentially requires libraries to change their definition of archival, and their understanding of the commitments and resources necessary to function as a repository.

#### *Preserving Context*

The issues involved in preserving context are conceptually the same for digital and traditional formats, but it is important to be aware of the new options introduced by characteristics of digital information. The concept of fixity in selecting and maintaining a version of a digital resource is related to understanding its context. How and why has a version been selected, and how does the preserved version relate to those that no longer remain? Provenance is important contextual information; the provenance

of published books can be adequately documented by publication information, but digital resources have no such conventions. The history of reproduction and migration is essential for digital objects that document phenomena, such as reproductions of artistic images, scientific data, etc., as the layers of translation and transformation for digital resources can be enormously complex. Maintaining the original “look and feel” of a digital resource, or at least documenting it for the future, will be increasingly challenging as dynamic documents must be sampled for preservation and as objects migrate to new platforms.

#### *Preserving Access*

There is little use in preserving a resource unless those who need it can easily discover its existence and make use of it. Network discovery and retrieval tools may make cataloging and indexing of digital resources easier, but it is the responsibility of a digital repository to ensure that access is adequate. Once a resource is discovered by a potential user, an infrastructure for delivery and use is required. Thus, a repository must maintain adequate network connectivity; software for retrieving, viewing, and if appropriate, manipulating digital information; a reliable system for referencing and locating the resources it holds; and systems for protecting the security and integrity of data it holds and for implementing any restrictions placed on access by the owners of the intellectual property. Since these access requirements are not specific to long term preservation, but are necessary as well for short term use of digital information, there is a large community of publishers and information providers interested in developing economical techniques for all of these functions. Libraries and other repositories will not be alone in their efforts to maintain access.

#### *How Do We Organize Ourselves to Preserve Digital Information?*

As there are so many kinds of individuals and organizations interested in providing access to digital information, the roles of information providers *vs* those of libraries and other repositories are not clearly defined in the new environment. In the world of material information, the distinction between preservation and use is quite clear. The more an object is used, the more it is subjected to wear and tear; use and reservation are not only different activities, they may be antithetical to each other. In this world, only libraries and archives have become concerned with long term preservation.

The boundaries between roles in the digital environment are, at present, less clear. Use keeps digital objects alive. If they are in demand,

the activities that enable their use — e.g. access, systems engineering, refreshing — are the same activities that will ensure their continued functionality. Digital objects that sit unused will become unusable. These blurred boundaries between immediate and long-term use have led many information providers to view themselves as the archival keepers of their data. They see no need for intermediaries, such as libraries, to provide archival functions. Their interest in maintaining archival control is related to their interest in maintaining intellectual property control over their information. As noted earlier, the focus of digital preservation is on content, not objects, and content owners do not wish to lose control over their property.

In an ideal environment, content developers, publishers and libraries would work closely together in developing interdependent roles. Those who create digital information would design digital resources with access and long term use in mind, e.g. by providing metadata, using standard formats, documenting software, etc. Those who publish digital information would deposit it with appropriate repositories and develop agreements for long-term preservation and access. We need to strive for this ideal environment. However, there will be cases where digital information has not been created and deposited with such foresight, and libraries will need the legal basis to acquire and aggressively rescue files that have been essentially abandoned by those who created or published them.

Our greatest challenges are organizational rather than technical, as the final report of the Task Force notes. We currently lack the infrastructure of practices, standards and organizations that is needed to support preservation of digital information. Elements of the infrastructure that we need to begin to build include the following:

- Legal bases for deposit and rescue. In individual countries and internationally, legislation and agreements are needed to encourage legal deposit of electronic resources with archival repositories, to enable rescue of abandoned resources, and to facilitate access and use of archival files.
- Standards for description. Current library cataloging standards are not sufficient to describe access and contextual information about digital resources.
- Standards for design and formats. Migration on a broad scale is only feasible if standard formats and platforms are widely used.
- Backward compatibility. Software manufacturers need to be educated and encouraged in the importance of maintaining the usability of older versions of their products.

- Accepted best practices for systems engineering and migration. Information about these techniques is not widely shared across professional communities.
- Enabling technologies. Functions such as migration, emulation and access can be facilitated by new systems design and technological development focused on these issues.
- Guidelines for archival practices and principles. The expected operational requirements to serve as a responsible repository for digital information are not well understood. Guidelines need to be promulgated, perhaps even through a certification process, so that organizations can develop themselves to meet the needs.
- Processing centers. Many libraries and archives will not have the technical capabilities or specialized expertise to maintain and migrate digital files, even though they may have the appropriate mission and skills to identify and take responsibility for files to be preserved. In a manner similar to vendors that provide services for microfilm production and storage, processing centers could provide operational services for digital preservation.
- Models for co-operative arrangements. Digital preservation is complex and expensive; it can only be practically undertaken on a distributed basis. Co-operative agreements will be essential to future use of digital files.

*Recommendations by the Task Force for Future Action*

The Task Force on the Archiving of Digital Information made recommendations in three areas: best practices; new models and techniques; and support structures.

*Best Practices.* Although digital resources are relatively new in the long history of communication media, they have been with us for some thirty years, and many have already been preserved. This work has been done in specialized sectors, such as social science data repositories and scientific observatories, and the practices used in each sector have not been shared across disciplines. As libraries begin to face the preservation of digital files created by mass communication and electronic publishing, they can learn from practices elsewhere. The Task Force recommended that the Commission on Preservation and Access commission case studies of successful work in areas such as: design of digital files, mass storage, resource description and migration paths.

*New Models and Techniques.* The best way to gain experience and learn about new approaches is to encourage and fund pilot projects. The Task

Force identified three areas in which funding programs should be developed to stimulate projects: (1) Co-operative efforts to rescue files. (2) Sponsored digital archives. (3) Research and development in new technologies.

*Support Structures.* Developing the needed organizational and societal infrastructure for digital preservation will require strategic initiatives in several areas. The Task Force recommended that the Commission on Preservation and Access and the Research Libraries Group develop implementation plans to achieve the following objectives:

- (1) Make preservation an explicit goal of the US Federal Government National Information Infrastructure initiative.
- (2) Articulate and lobby for the legal principles that will facilitate digital preservation.
- (3) Develop criteria for organizations that wish to serve as digital repositories.
- (4) Educate scholarly societies about the importance of digital preservation and engage them in preserving the information they produce.
- (5) Maintain international coordination.

The issues and strategies identified by the Task Force are not specific to the US. The challenges of preserving digital resources are large and complex; meeting them will require many approaches and many perspectives. Strong, active national libraries across the world have a special role to play, as do research libraries, archives and a variety of specialized repositories. We will need to share information about best practices and new technologies across a wide spectrum of communities that have not worked together in the past. Digital preservation has emerged as a new, critically important field of interdisciplinary and international activity, and much work is ahead of all of us to find answers to the many questions.

This is a prime topic for international co-operation. Many in the US are unconvinced that publishers will ever agree to either of the two requirements for digital preservation — certified archives and a failsafe mechanism — that are outlined by the Task Force. If these preconditions are unacceptable, what are the alternatives? As professionals, we must seek common ground without international colleagues and proceed to specify what will be necessary to assure access to our cultural and intellectual heritages for as far into the future as possible. Preservation is a role that has been entrusted to us by society. Digital technology makes it possible to provide new and exciting methods of access to information, but in the



process we cannot abdicate our responsibility for preservation—the embalming of dead genius—so that others will be able to build on the knowledge created in all formats.