Shifting from paper to Digital Records - Preservation?*

Michele V. Cloonan**

◁ 목 차 ▷

<국문초록>

전자문서의 보존은 보존관리인들에게 새로운 과제를 가져다 주었으며, 오늘날처럼 정보를 파악하기 힘든 적은 없었다. 문서가 일단 매체로 전환되면 수 백년 혹은 수 천년까지 영속할 수 있었으나, 전자문서는 사라질 위기에 있어서 물리적으로는 이용할 수 없고 법적으로는 적합하지 않아 받아들여질 수 없는데, 이러한 현상이 나타나는 원인은 매체가 더 이상 쓰이지 않는다든가, 기술력이 시대에 뒤떨어져 쓰이지 않는다든가, 표준이나 안내지침이 부족하거나, 보존관리인들의 전자문서관리와 보존을 위한 계획안을 마련하지 못한다는 등의 이유 때문이라 할 수 있다. 종이자료와 달리, 전자매체는 물리적인 객체로 보존될 수 없고, 전자문서는 단지 문서를 재생할 능력을 보존할 뿐이다. 그리하여 전자문서의 적절한 관리가 문서보존에 더할 수 없이 중요하게 되었으며 전자정보관리의 초창기인 지금 우리가 최신의 문화유산이 지닌 잠재력과 함정을 모두 이해하기는 힘들기 때문에 이를 완전히 이해할 때까지 완전한 보존해야 할 것이다.

I. The Digital Present

The preservation of electronic records presents new challenges for records managers:

* 이 논문은 "제3회 기록관리학 국제학술대회 / 기록매체의 보존관리"에서 발표된 논문임.

** Associate Professor, UCLA, USA

information has never before been as fugitive as it is today. Electronic records are in danger of disappearing, becoming physically unusable or legally inadmissible, almost immediately. There are many causes for the short life span of digital and electronic records: media deterioration, technological obsolescence, a paucity of standards and guidelines, but also key is the failure of many managers to plan for the maintenance and preservation of electronic records.

In a recent review of electronic archiving issues and problems, Dorothy Warner expresses concern about the lack of preservation plans for digital government records at either the state or federal government levels.[1] At the federal level "there is no overall plan for archiving federal government records that exist only in digital format. Instead each agency determines its own preservation policy."[2] A failure top plan for preservation exacerbates already complex digital preservation issues-not all problems are technological.

The goal of preserving inherently unstable media may at first appear to be a contradiction. The dictionary provides two definitions of preserve:

1. To maintain safety from injury, peril or harm; Protect; and
2. To keep in perfect or unaltered condition; maintain unchanged.[3]

We can put safe guards in place to protect both paper-based and digital media, but how can we maintain digital material "unchanged?" We can't, because while paper is an essentially static medium, digital media is dynamic and constantly changing.

How are today's archivists approaching the preservation of electronic records and digital media?[4] Archivists need to ensure the authenticity, reliability, and long-term

1) Dorothy Warner,"'Why Do We Need to Keep This in Print? It's on the Web...': A Review of Electronic Archiving Issues and Problems". Progressive Librarian 19-20 (Spring 2002): 1.
2) Warner, p. 1.
3) The American Heritage Dictionary of the English Language. 3rd. ed. Boston: Houghton Mifflin, 1992).
4) Records created or maintained in electronic form, either analog or digital, are herein referred to as electronic records. Digital preservation refers more broadly to reformatted items, born-digital

accessibility of permanent electronic records for current and subsequent users.[5] Traditionally they have done so by gathering documents, establishing provenance, and maintaining and demonstrating an unbroken chain of custody in an evidence-based approach to managing records.[6] Is it possible to ensure authenticity and reliability of records regardless of their formats? If so, which technologies and techniques are archivists using to preserve electronic records at the time of their creation as well as throughout their life cycle?

II. The InterPARES study

To learn more about current archival approaches to digital preservation, Shelby Sanett and I surveyed the activities of thirteen institutions and research projects in the United states, Canada, Europe, and Australia that employ or are exploring strategies to preserve authentic electronic records.[7] [We carried out this research as part of our work on the preservation Task Force of the InterPARES Project-International Research on permanent Authentic Records in Electronic Systems.]

---

electronic materials, and born-again digital materials. We use the terms interchangeably. Recently, the term long-term retention seems to be gaining currency; it is used broadly to refer to be electronic records and digital preservation.

5) For various definitions of authenticity and authenticity of digital information, see Nancy Brodie, "Authenticity, Preservation and Access in Digital Collections," The New Review of Academic Librarianship 6 (2000): 225-238.

6) Anne Gilliand-Swetland, Enduring Paradigms, New Opportunities: The Value of the Archival Perspective in the Digital Environment. (Washington D.C. : Council on Library and Information Resources, February 2000): 11-12.

7) We carried out this research as part of our work on the Preservation Task Force of the InterPARES Projects (International Research on Permanent Authentic Records in Electronic Systems). InterPAREs is an international research initiative that involves national archives, university archives, and a team of academic researchers in archival science, preservation, and computer science to address issues related to the permanent preservation of authentic electronic records.

These strategies include preservation techniques (e.g. refreshing migration, emulation); selection for preservation; staffing configurations; cost modeling; access to preserved records; and policy making. We carried out the survey twice: in 2000-01 and in 2001-02. A lengthy questionnaire, divided into fourteen sections, was mailed to each participant. The sections covered the following topics:

A) Information about the Institution Project
B) Program and policy
C) Specifics of Preservation Technique/Method/strategy
D) Selection for preservation
E) Staffing,
G) Technical Questions
H) Costs
I) Preserving Records
I) Description/Documentation of Preservation processes
K) Access to preserved Records
L) Charges
M) Reproduction and Copyright
N) Preservation Policies.

After we received the responses, we interviewed the respondents on the telephone in order to receive clarification on individual questions and to gain further insights about the programs. I will briefly summarize our findings, then I will discuss some U. S. projects currently under way.

Ⅲ. The Findings[8)]

8) For the full findings of Round 1 of the Study, see our "Preservation Strategies for Electronic Records: Where We Are Now: Obliquity and Squint?" American Archivist, forthcoming Spring/Summer 2002.

The interviews indicate three broad themes. First, that the perception of what preservation is goes beyond archival practice tome media being preserved. Because electronic material is inherently ephemeral, and the time frame involved to preserve and provide access to this material extends to perpetuity, the traditional definitions of preservation may not apply. Indeed, a shift is already apparent, from defining preservation as a once-and-forever approach for paper based materials, to an all-the-time approach for digital materials By "once-and-forever" we mean that specific actions can be taken to insure longevity in paper-based materials such as proper environmental and storage facilities, proper housing security measures, etc. When such measures are in place, paper materials can last indefinitely barring disaster, infestation, or vandalism. While paper-based materials also require constant care, benign neglect is not always reformatting, migrating, etc., which is a much more pro-active and costly endeavor.

Preservation managers must safeguard electronic records against the threat of interrupted management. And, most important, the purpose of preservation is to authenticate documents so that they are reliable and trust worthy. For digital materials, neglect may result in unreliable documents or, worse yet, in total loss.

Second, the rush to develop the technological processes necessary to preserve authentic electronic records appears to be at the expense of directly addressing cost and policy issues at the start of preservation projects. One said, "We haven't yet been asked to measure costs!" We don't need to justify costs. Fixed costs are unknown." The problems posed by preserving authentic electronic records permanently (or for as long as possible), requires the development of a cost model, which will be unique and not a hybrid of exciting digital conversion cost models.[9] Preservation must begin with the creation of electronic material.

A cost model for preserving electronic records will need to reflect this perspective, which differs from the traditional preservation point of other view. Costs, however,

9) Shelby Sanett, forthcoming in College & Research Libraries (July 2002).

must not single-handedly be used to justify not preserving otherwise valuable records.

Last, the lack of preservation policies in place is a distinct gap in the research design of many of the projects and possibly reflects a lack of commitment among the stakeholders in institutions. What is the reasoning behind developing policy as an end result of a projects, instead of concurrently with it's progress?

Meeting the technological challenges of preserving electronic records is more of a priority within these institutions than developing policy. Will the overall progress in this arena will be more uneven than is necessary because institutions are failing to develop policy? Several institutions that responded to our survey have had active programs for a long time and their policies evolved, rather than being strategically planned. It is practically impossible to set policy 100% at the outset of a project-especially one in such a complex areas as the preservation of electronic records. policy will naturally evolve rather rapidly at the outset of the program when the practitioners encounter new, possibly unanticipated features of the program that require policy decisions. As the program matures, and even while it is still developing policy will concomitantly need to be re-thought or new conceived. In fact, policy must also drive technological development. When the program is in "full swing, policy win have reached a point at which it is now well thought out, though still subject to modification, as the program requires.

In subsequent research, we hope to explore not only the "why",behind the positioning of policy development within the institution, but also the development of its content. We want to explore the roll of the stakeholders and the influence of the legal and political environments that provide the context in which policy is formed.

Ⅳ. Other InterPARES Approaches to the preservation of Electronic Records

The goal of InterPARES is to develop the theoretical and methodological knowledge essential for the permanent preservation of records generated electronically, and, on the basis of that knowledge, to formulate model ploicies, strategies, and standards capable of ensuring their preservation."[10] In addition to the preservation survey, the InterPARES Preservation Task developed s Model for preserving Electronic Records. The objective of the model is to provide a framework for preserving electronic records. The approach was to use the IDEF-0 notation and methodological to represent both the problem and the results of the Task Force's analysis of the problem. Further, this model is compatible with the open Archival Information system (OAIS) reference model maintained by the International organization for standardization (ISO), and supported by the RLG.[11] [The OAIS model was developed in the early 1990s by a subsidiary of NASA.] OAIS has become the model of choice for digital preservation.

V. Other Research Activities

I would like to briefly describe a few research activities currently underway in the United States in order to give an idea of me range of approaches that are currently being taken.

VI. NARA

NARA is currently working on a preservation strategy that will "create a digital

10) InterPARES Projects: http:/www.interpares.org
11) International Organization for Standardization (ISO). http://www.iso.org

National Archives that will make the invaluable record of America available electronically to any one, at any time, many place, for as long as needed."[12]

The Electronic Records Archives (ERA) program is currently involved in a large-scale Research & Development projects to validate this vision. Started in 19%, it is known as Persistent Object Preservation. partnerships from the core of ERA with the open Archival Information System Reference Model laying the foundation. The InterPARES project-with representatives from ten national archives-also forms a partnership. The National partnership for Advanced computational Infrastructure (NPACI), led by the San Diego Super Computer Center, is developing the requirements and processes for the preservation and reproduction of authentic records through the development of the persistent archives method which is essentially a form of migration using XML Document Type Definitions (DTDS). There are three parts persistent object preservation; Ingest, Management, and Access.

Ⅶ. National Digital Information Infrastructure and Preservation Program(NDIIPP)

In December 2000, congress passed legislation establishing the NDIIPP (PL 106-554) in the Library of congress. In it, LC is charged with leading the planning effort for the long-term preservation of digital works. LC is working with federal, research, and library organizations, as well as business. The legislation allocated $100 million for the program, to be released in stages.[13]

LC is still in the early stages of this program. In 2001 it held three stakeholder meetings; so far there is widespread support for a national initiative for the long-tenn preservation of digital content.

12) NARA Electronic Records Archives (ERA) Program. http://www.nara.gov/era

13) Amy Friedlander, "The National Digital Infrastructure Preservation program: Expectations, Realities, Choices, and Progress to Date." D-Lib Magazine 8.4 (April 2002): 1.

Ⅷ. Find-It! Illinois Program

There are a number of stakeholders who need to play an active role in digital preservation including-but not limited to-archivists, librarians, authors, publishers, distributors, networked information providers, IT suppliers, universities, government, legal depositories, consortia, and research funders. Find-It! is "an example of a partnership that attempts to recognize all of the stakeholders. The Find-It! Illinois Program is one of the state-level Government Information Locator Service (GILS) programs around the United States."[14] One such program is its uniform metadata tagging, which was accomplished through liaisons between state agency librarians and their agency webmasters. A tool provided by the Illinois state Library, the Metadata Generator, faccilitates the process.[15] The Library also provides controlled vocabulary through their "Jessica Tree." Thiss subject hierarchy enables interstate government information access.

Ⅸ. The Internet Archive

This is an archive or "snapshots" taken of selected web pages by Alexa, an information company. In collaboration with the Library of Congess, the Internet Archive culled data from multiple sources on September 11m and in the weeks immediately following.[16]

14) Warner, p. 9.
15) http://www.finditllinois.org/metadata/webmasters.htm
16) http://www.archive.org/about

X. Lots of copies Keep stuff safe (LOCKSS)

This program is part of the Andrew W. Mellon Foundations e-journal archiving program in which seven libraries are participating: the New York Public Library and the university libraries of cornell, Harvard, MIT, Pennsylvania, Yale, and Stanford. Stanford is developing archiving software tools under its LOCKSS program.

XI. The Research Libraries Group (RLG) and The online computer Library center (OCLC)

RLG and OCLC are jointly conducting preservation research. Thus far they have focussed on No areas: preservation metadata and the attributes of a digital archival repository.

XII. Conclusions

It is possible to develop models, protocols, and standards that factor in such problems as media deterioration and technological obsolescence, and work is being done in these areas. For example, some institutions already safeguard their digital files creating backups and storing them offsite, and through routine migration. Others maintain metadata separately from their master files of electronic records. The implementation of more standards may satisfy the basic requirements for preservation: to maintain authentic and reliable records for as long as they are needed.

Yet attention must also be paid to "the preservation of the information stored in those

media during recurrent migration to new software and hardware. In the process, many of the intrinsic characteristics of information objects can disappear-data structures can be modified and presentation of the object on a computer screen can be altered."[17]

How do we preserve digital media? Unlike paper objects, digital media cannot be preserved as physical objects. In stead, for electronic records, we can only preserve the ability to reproduce documents. Therefore, the proper management of electronic records is crucial to the preservation of them. In this early stage of electronic information, it is difficult to understand all the potential or all of the pitfalls of our newest cultural heritage. Until we do, however, there will be no preservation of it. Archivists understand the life cycle of documents, from data creation, to structure, description, documentation, storage, management, and use. By incorporating preservation into each stage of the life cycle, archivists are best positioned to assure the preservation of electronic records for as long as they are needed.

17) Gilliland-Swetland, 13; this issue was also raised by several of our respondents.