

Preservation Strategies for Electronic Records: Where We Are Now—Obliquity and Squint?

Michèle V. Cloonan and Shelby Sanett

Abstract

This article reports on a survey of the activities of thirteen institutions and projects in the United States and abroad that employ or are exploring strategies to preserve 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 policymaking. Particular attention is paid to three broad areas: the evolution of the definition of “preservation,” the role of costing in preserving electronic records, and the gap in policy development in which to situate and strategize the present and future preservation of electronic records. By documenting the variety of approaches that are being taken, the authors seek not only to shed light on current practices, but also to offer informed consideration on where preservation might be headed.

Information has never before been as fugitive as it is today. Whereas records were once written on media that could last hundreds—or even thousands—of years, 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, and the failure of many managers to plan for the maintenance and preservation of electronic records. Howard Besser refers to this last cause as “the custodial problem.”¹

This study was prepared, in part, with support from the National Historical Publications and Records Commission (NHPRC), grant number 99-073.

¹ Howard Besser, “Digital Longevity,” in *Handbook for Digital Projects: A Management Tool for Preservation and Access*. (Andover, Mass: Northeast Document Conservation Center, 2000), 156.

Archivists need to ensure the authenticity, reliability, and long-term accessibility of permanent electronic records for current and subsequent users.² 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.³ But the advent of electronic records has raised many questions about their long-term preservation. 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? How are archivists defining the term *preservation*? Do varying definitions of the term reflect differences in approaches?

Preservation of electronic records is a young field. Indeed, some people believe that the very concept of electronic records preservation is an oxymoron. The term *preservation* implies permanence, yet such media are inherently unstable. Adding to the ongoing problem of technological obsolescence, such challenges as those posed by issues of copyright and fair use, organizational and custodial questions, the concerns and interests of different stakeholders, substantial financial resources required for preservation programs, and legal admissibility requirements, make the prospects for preservation of electronic records seem grim indeed.

Purpose of the Study

The purpose of this study is to identify and describe the strategies and techniques that are in use or in development, in institutions and in research projects, to preserve electronic records. Knowledge of these efforts should help inform archivists as they formulate strategies, policies, and standards for preservation.

An opportunity to explore these issues presented itself through our association with the InterPARES Project (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 in addressing issues related to the permanent preservation of authentic electronic records. The project is investigating and developing theoretical frameworks, methodologies, and prototype systems required for the permanent

²For various definitions of *authenticity* and *authenticity of digital information*, see Nancy Brodie, "Authenticity, Preservation and Access in Digital Collections," *New Review of Academic Librarianship* 6 (2000): 225–38.

³Anne Gilliland-Swetland, *Enduring Paradigm, 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.

preservation of authentic electronic records. The research reported in this article was part of our work on the InterPARES Preservation Task Force. Its purpose was to explore one of the five research questions that the Task Force was charged with examining: What methods, procedures, and rules of long-term preservation are in use or being developed?

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

Counterintuitively, perhaps, it is during the preservation of digital materials that evidential value is often most at risk of being compromised. Digital preservation techniques have moved beyond a concern for the longevity of digital media to a concern for 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.⁴

Since the field of electronic records preservation is a fledgling one, with many concepts still to be comprehended and issues resolved, we decided to conduct this study in three rounds, from 2000 through 2003. Round one (2000–2001) surveyed thirteen institutions, programs, and projects. This round also provides a baseline of data for the second and third rounds. In round two (2001–2002), we are administering a revised version of the questionnaire and interviewing key informants. In the third round (2002–2003) we will conduct case studies on an even smaller number of programs and projects. By the end of the three rounds, we expect to be able to present a sharper picture of preservation strategies in practice than currently exists. These strategies include preservation techniques (e.g., refreshing, migration, emulation); selection for preservation; staffing configurations; cost modeling; access to preserved records; and policymaking.

Although other studies have explored individual aspects of electronic and digital preservation—such as standards, intellectual property, or specific techniques such as emulation—we are examining the continuum of activities that constitute the broad range and long-term goals of preserving electronic records. By documenting the variety of approaches that are being taken, we

⁴Gilliland-Swetland, *Enduring Paradigm, New Opportunities*, 13.

hope not only to shed light on current practices, but by the end of the third round, to offer informed insights into where the field might be headed.

When we began our research, no study had yet provided a comparison of preservation techniques for electronic records across institutions and projects. The closest research to ours is a study that was carried out by Margaret Hedstrom and Sheon Montgomery to examine long-term retention needs and requirements in the Research Libraries Group (RLG) member institutions.⁵ Another related work is the book on digital imaging for libraries and archives edited by Anne Kenney and Oya Rieger.⁶ One section of this book, “Digital Preservation Strategies,” gives a succinct overview of techniques such as refreshing, migration, and emulation. We designed this study to examine specific long-term retention strategies in more detail.

Research Questions

As indicated previously, our initial set of questions was formulated by the InterPARES Preservation Task Force:

- What methods, procedures, and rules of long-term preservation are in use or being developed?
- a. Which of these meet the conceptual requirements for authenticity?
 - b. Which methods of long-term preservation need to be developed?
 - c. Which of these methods are required (or subject to standards, regulations, and guidelines) in specific industry or institutional settings?

To this we added three more sets of questions:

- What is the meaning of *preservation*? Does the meaning change when it is applied to electronic rather than paper-based records?
- Will current strategies for preserving electronic records ensure longevity and authenticity?
- How are costs for the preservation of electronic records derived? Have effective cost models been developed?

Our three-round study seeks to ascertain the strategies that are either currently in use or are in development for preserving electronic records. We also plan to track the evaluative studies of these techniques.

⁵Margaret Hedstrom and Sheon Montgomery, *Digital Preservation Needs and Requirements in RLG Member Institutions*. (Mountain View, Calif: Research Libraries Group, 1998) <<http://www.rlg.org/preserv/digpres.html>> (May 9, 2002).

⁶*Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, Calif.: Research Libraries Group, 2000). The contributors cover particular topics (e.g. “What Users Want from Digital Image Collections,” “Benchmarking for Conversion,” “What About Copyright?”), but the book does not compare practices across institutions.

Definitions of Terms

The terminology used in discussing the preservation of electronic records can be problematic. Records created or maintained in electronic form, either analog or digital, are herein referred to as “electronic records.” It is impossible to avoid using the term “digital preservation,” however, because that is the term most frequently used in the preservation literature. (For example, in the works cited here, nearly every author uses it.) We recognize that it is not always the most precise term for archivists because it refers to the preservation of reformatted items, born-digital electronic materials, and born-again digital materials. We will use the term “digital preservation” to reflect the general preservation discourse found both in the literature and in our interviews, as well as “preservation of electronic records,” when appropriate. Recently, the term “long-term retention” seems to be gaining currency.⁷ As this term is more general than either “digital preservation” or “preservation of electronic records,” we will also use it where appropriate. This flux in the terminology is, we believe, reflective of an evolving new field.

The definitions that we used for this study are in appendix 1. We developed our own working definitions based on the literature, responses on the questionnaires, and/or our own knowledge of the field. We hope that these definitions—where appropriate—can become the standard terminology of the field.

Research Methodology

To answer our overarching questions, we decided to use a purposive sampling strategy, one that would show different perspectives on the problems we wanted to address. We surveyed a sample of archives, projects, and programs in the United States, Australia, Canada, and Europe. We chose to collect data using a questionnaire developed with feedback from other members of the InterPARES Preservation Task Force.

We identified fifteen sites known to be developing one or more of the following preservation techniques: refreshing, migration, emulation, collection-based persistent object preservation, bundling, the Universal Preservation Format, and robotics. We ultimately interviewed representatives from thirteen of the fifteen sites that we selected. (A list of sites is provided in Appendix 2.) Such a small population did not warrant a quantitative research design. Further, since we knew that we would be asking broad, sometimes open-ended questions, we decided on a qualitative design. Since it was not our goal to make statistical inferences, but to learn about processes and methodologies, we explored one research method and one research strategy: the case study and the survey.

⁷See, for example, *Research Libraries Group News* 52 (Spring 2001): 3, 8–9.

We ruled out the case study method for round one because we felt that it was too early in the development of long-term retention strategies to study individual programs in depth. Instead, this methodology will be used in round three. Rather, by using a questionnaire for round one, we hoped to establish a benchmark of current practices from which we could collect a general comprehensive view.⁸ The questionnaire would provide baseline data about the current (year 2000) state of long-term retention practices that we could draw on in the subsequent rounds of the research.

The participants in the study were sent a consent letter, which explained that if they volunteered to participate in this study, they agreed to read over the attached survey instrument and participate in a telephone interview based on this instrument. (A copy of the questionnaire is included in appendix 3 of this article.) We followed up on this letter by telephone or e-mail to arrange a suitable time for the interview. The telephone interview was not taped. In a few instances, we were able to conduct the interviews in person. As part of our consent agreement with participants, we agreed not to disseminate any proprietary information, quote any of the interviewees, or disclose individual or institutional identities without express written permission.

Boundaries and Limitations of the Study

The data gathered and analyzed in round one of this research allow us to draw only tentative conclusions about current preservation techniques in a limited number of venues (six archives, six programs and research projects, and one library). Although our aim was to identify as many different preservation techniques as possible—without regard to how many institutions and projects were experimenting with new techniques—there is the possibility that we missed learning about important new projects. Also, preservation techniques such as bundling and the Universal Preservation Format (UPF) are still only in the earliest stages of development, so we did not learn as much about them as we had hoped. However, the study has the advantage of identifying and describing cutting-edge approaches to the preservation of electronic records, and we will continue to monitor new developments.

Another limitation is that InterPARES sponsored this research, and many of the participants in our study were affiliated with the project. Therefore, the needs and perspectives of the InterPARES Preservation Task Force drove the study. Since InterPARES focuses on electronic *records*, we focused on archives and archival projects. We selected one library for the study, but the practices of that library demonstrated it to be a disconfirming case and we will not include

⁸We based some of our questions on those used in: IFLA/UNESCO, *Survey on Digitisation and Preservation*. (Wetherby, UK: International Federation of Library Associations and Institutions, 1999).

it in round two. On the other hand, InterPARES is an international project with great visibility. Our affiliation with the project may have resulted in entrées to research projects that we might not have otherwise been able to gain access to.

The data from our interviews with participants are probably not as comprehensive as they might have been had we taped them. We chose not to tape the interviewees, thinking that it would cause the participants, as representatives of their institutions, programs, or projects, to be constrained in their responses. During the interviews, both researchers took notes. After the interviews were completed, and upon further review of our notes and procedures, we felt that our notes represented more of an interpretive record, rather than transcripts. The participants did not appear to be constrained in their responses, and this observation led us to consider the possibility of taping subsequent interviews. An informal survey of some of the participants showed that they would not object to having their interviews taped for the second round of the survey. We feel that using the telephone during the interviews also introduced a barrier between the respondents and the researchers. This was especially apparent in interviews where respondents spoke English as a second or third language. In the second round of the preservation survey currently underway, we have asked for electronic responses to the survey instrument, and the follow-up telephone interviews are being taped.

A final potential limitation of this study relates to concepts and terminology, a limitation pointed out by Hedstrom and Montgomery in their RLG study.⁹ We also surveyed practitioners and researchers in both the United States and abroad. As the RLG study suggests, we found that differences in terminology and concepts may reflect different cultural perspectives; or, since our interviewees included archivists, librarians, computer scientists, and engineers, professional differences.

Survey Results

The questionnaire was divided into 14 sections: A.) Information about the Institution/Project and Respondent, B.) Program and Policy, C.) Specifics of Preservation Technique/Method/Strategy, D.) Selection for Preservation, E.) Cooperation, F.) Staffing, G.) Technical Questions, H.) Costs, I.) Preserving Records, J.) Description/Documentation of Preservation Processes, K.) Access to Preserved Records, L.) Charges, M.) Reproduction and Copyright, and N.) Preservation Policies.

Not all sections or questions were relevant to each project or program. However, we chose to be comprehensive in order to learn about as many aspects of each program as possible. The following discussion of the survey uses the

⁹Hedstrom and Montgomery, *Digital Preservation Needs and Requirements*, 5.

numbers 1 through 12, to refer to the survey respondents. The tabulated responses appear with the full questionnaire in appendix 3. This portion of the article will summarize the results in each section of the questionnaire. Three key areas, however, receive extended discussion in the subsequent portion of the article.

**Section A: Information about the Institution/Project
and Respondent**

This section is summarized in appendices 2 and 3.

Section B: Program and Policy

The purpose of this section was to ascertain the range of preservation activities for electronic records, and it began by asking each respondent to give an overview of his or her program. When asked, “When did your institution’s program or activities begin?” respondents traced electronic preservation programs back to the inception of *any* preservation activity in their institutions. Four respondents reported that their programs started in the 1970s, one in the 1980s, and five in the 1990s. Of the four who reported activity since the 1970s, two archives reported that they had been preserving digital objects since the 1970s. Three participants responded “not applicable” because they represented special projects. Each institutional respondent then described the development of preservation programs over time.

Respondents were asked to describe institutional issues that have an impact on their programs. The issues mentioned included: the problems associated with outsourcing; inadequate staffing; storage; prioritization of records to be preserved; inadequate resources, including funding; and increasing legal mandates for preservation.

We also asked what methods or techniques they were exploring or using for digital preservation. Four respondents said migration, and three said knowledge-based persistent object preservation. Emulation and physical preservation were cited by two respondents each. Bundling, refreshing, digital archaeology, preservation copying, and robotics were each cited by one respondent. It is important to note that some institutions and projects are using more than one strategy, and that bundling is in the exploratory stages only.

The next question asked respondents to describe the items that they are preserving. Respondent 1 reported preserving only born-electronic records. Others preserved a variety of born-electronic records and other digital artifacts including spreadsheets, databases, computer games, geomatics, and serials.

One question in this section caused confusion for some respondents: “Do you consider any of these materials to be records?” Six answered “yes,” one

answered “no” and 6 answered “not applicable.” Not all of the respondents distinguished between records, documents, and items. Our follow-up question, “Do you make any special provisions for preserving records, as opposed to other types of digital materials? Only respondent 11 said “yes.” Five said “no” and seven responded “not applicable,” perhaps because of the same confusion demonstrated in the preceding question. We also asked if the program or project was testing or evaluating any of the methods or techniques currently being used. Only three respondents are engaged in testing or evaluation.

The next question asked respondents to identify problems, difficulties, or threats to the integrity of digital materials with any of the above-mentioned techniques. Three respondents identified preserving the integrity of the original digital object as a threat. Respondent 4 discussed the issue of “acceptable loss” due to migration. Two other respondents identified the problem of changing standards while respondent 12 mentioned technology obsolescence.

From this section of the questionnaire, it became apparent that most of the respondents were better versed in identifying problems than in developing solutions. The final question asked respondents to define preservation, and these responses are described in detail below.

Section C: Specifics of Preservation Technique/Method or Strategy

This section asked respondents to elaborate on the preservation techniques that they identified in Section B. As reported above, the most common strategy is migration. Respondents were asked in this section to tell us how they selected the preservation method employed. Respondent 1 identified migration as the standard method of moving from one platform to another. Respondent 7 responded that her institution believes it to be the best method.

As a follow up question, we asked if any of the institutions or projects used a hybrid approach that combined two or more techniques. Four respondents said “yes,” four said “not applicable,” one said “maybe,” and three said “no.” Of the four “yes” answers, three indicated microfilming and scanning and one said scanning and physical treatment. No one reported combining such techniques as robotics and migration, for example.

We also asked whether the respondents had considered what effect their chosen techniques might have on the intellectual integrity of the digital materials. Nine respondents indicated “yes;” four, “no.” Respondent 7 criticized the work of an institution (one not included in our study) that had enhanced original photos, by citing issues of veracity. Respondent 8 worried that emulation may not work for some classes of materials.

The final question in this section asked whether there was evaluative data on the efficacy of the preservation method/model used by each respondent.

Only two respondents had formal evaluation procedures. Respondent 1 reported that audit trails are kept, and that his institution was moving to a document management program to support audit trail development. Respondent 11 uses Knowledge-Based Persistent Object Preservation that conforms to the Open Archival Information System (OAIS) model.

Section D: Selection for Preservation

In this section we asked respondents to identify the criteria they used in selecting which materials to preserve, as well as where the selected materials came from. There were thirteen responses, including one “not applicable.” *Historical/cultural value* and *Legal requirement to preserve* were the two criteria guiding selection of materials for preservation for nine of the respondents. The next two criteria were “research into preservation” and “retarding deterioration” (five respondents). “Saving space” and “Other” ranked third. No one cited commercial reasons as criteria for selection. Criteria mentioned in the “Other” category included: materials publishers wanted them to have; institutional requirements; asset management considerations; sampling Internet sites; and supporting the curriculum. The “not applicable” response was from a respondent who preserves materials for clients.

The materials selected for preservation, came from: government agencies (6); parent institutions (5); other, e.g. private, commercial publishers, politician’s private papers, private individuals, commercial entities (3); and other organizations or associations, e.g. corporations (private sector) (2).

Section E: Cooperation

Of the thirteen respondents, twelve cooperate with other organizations in developing their preservation program, while one does not. The types of collaborating organizations include archives (9), libraries (9), public companies (5), museums (3); and, in the “other” category, government agencies, other programs and projects, for profit and non-profit educational institutions, and universities. Cooperation occurred on an international level (11), national level (9), local (2) and with shared facilities (1). Work was distributed both equally (5); and in a different way (6).

Collaborations tended to evolve rather than to be planned. Participants viewed the strengths of collaboration as being in the sharing of responsibilities and costs; shared information and resources; and the opportunities to develop a consensual outlook. Weaknesses of the collaborative process noted include: resource costs; the problem of people being far apart and influenced by local interests and resources; span of control issues; and the tendency for research and development to go off on non-productive tangents.

Section F: Staffing

The responses indicate that the majority of electronic preservation staff members are part time, work under the supervision of a person who usually holds at least one master's degree, and divide their time with other projects and departments. Only two of the positions were held by Ph.D.'s, and both of these individuals worked on research projects. Educational and professional backgrounds and skills include: computer science, archives, libraries, people with management and history backgrounds, and people who learned their technical skills on the job. One of the individuals has a Ph.D. in history and preservation and directs a facility.

Section G: Technical Questions

Ten respondents carry out preservation in-house, while one uses a commercial vendor. One respondent uses both in-house and commercial resources, while another respondent has disks on a shelf and is not taking pro-active steps to preserve them at this time.

Pre-preservation preparation of records includes: documenting provenance; checking to make sure that records have not been tampered with; inspecting them; physical preparation as needed; putting the records into a standard format; accessioning; and arranging and describing the records prior to copying them.

The questionnaire asked respondents to discuss weak points in their institution's preservation methods or techniques. These include database problems. For example, some respondents working with GIS and CAD materials found that tabular displays don't work. For electronic documents, respondents cited compound records, web site material, and attachments and nested materials as being problems. Implementing finding aids for a million collections is a problem in one archives. Another archivist said that his archives has not yet done the right thing with respect to textual documents. According to him, there are not enough resources, i.e., money and people; and the current staff does not have sufficient expertise for the work. Lastly, an archivist of a large national archives stated that the archival profession has not yet articulated its needs regarding system requirements; that the profession "is not used to system thinking in work, and never had an opportunity to do it before." During round two, we will explore the scope of these issues in more detail in order to determine whether they are ongoing, and whether resolutions have been investigated or applied.

Quality control methods applied to the preservation process or activity include the use of a quality control manual for a large European archives, with more information in development. One respondent stated that his facility was currently using *ad hoc* methods not worthy of sharing. Another respondent cited quality control of metadata as being the most difficult technical challenge. One respondent stated that quality control methods would be a partial out-

come of their project. An archivist in a European archives stated that there was no really organized quality control; they rely on the professionalism of the individual, who often has no formal training. Other responses include the use of standard information technology techniques for verifying the quality or success of the copying and peer review research to promote publishing and discussion.

Permanent storage of electronic records includes: archival storage at partner sites; containers; server/redundant servers; and underground storage. One site is not storing backup copies of records offsite.

Section H: Costs

An extended discussion of this section is provided below.

Section I: Preserving Records

The majority of respondents who answered the questions stated that in their institutions, they organized records according to the standards currently in use, e.g., the same as for paper records; or according to the archival arrangement and description standards adopted by their institution. Five respondents stated that the question did not apply, while one institution responded that an organizational scheme would be an outcome of their project.

Regarding respecting provenance, four respondents stated that the question did not apply, while the balance of the respondents cited the development of metadata standards, describing records according to provenance and type of record, and conformance with the institution's general practice as examples.

Six respondents stated that the question regarding restrictions on access to records did not apply. However, all of the remaining respondents cited examples of the restrictions that their institution imposed on records. Respondent 1 stated that there was a thirty-year closure [from general public access] unless the record was in the public domain; and that the record can in fact, be closed for up to one hundred years. Respondent 4 also cited a thirty-year closure, adding that the records were still subject to subpoena. Other respondents stated that access to records is negotiated individually with each donor, while respondent 11 explained that the type of restriction on the record would determine accessibility.

Section J: Description/Documentation of Preservation Processes

Respondents generally described recordkeeping for the preserved material as detailed audit trails, catalogs, databases, part of corporate descriptive systems, or part of the metadata accompanying the preserved objects. The question did not apply to six respondents, because they are not yet using metadata. We expect to see more widespread use of preservation metadata by round three. (RLG and OCLC are collaborating on the international Preservation

Metadata Working Group, which is establishing approaches for preservation metadata that will work in a variety of settings and for a variety of materials.)

Nine institutions use recognized standards to describe preserved materials. These standards include Generalized International Standard Archival Description, Encoded Archival Description, MARC, modified Library of Congress Subject Headings, and the Dublin Core.

We asked whether the records for preserved materials and the original were the same or independent of each other, and most of our respondents did not understand the question. Eight respondents said the question did not apply, while two respondents stated that the record(s) were both the same and independent. Two respondents stated that the record was the same, while respondent 10 stated that the institution does not recognize that preserved materials are copies; both copies are the records.

According to the respondents, five institutions are addressing metadata use. The question did not apply to four other institutions. Respondent 6 stated that the use of metadata is extensive, e.g., content description, preservation history, but that it was not always completed. We take that to mean that the metadata record was not always produced in its entirety, according to institutional standards. Respondent 10 stated that metadata is being used in a variety of ways.

Section K: Access to Preserved Records

The majority of institutions make records available through a web site (7). One institution makes hard copy available, while one institution makes copyright protected material available to the academic community only. We did not ask whether the records were full text, or all or some series. This question did not apply to four institutions.

Question 12.2, about archival workstation connectivity, did not apply to twelve institutions, while the remaining respondent stated that the institution had dedicated machines that consisted of five workstations, some online, and some connected to secure servers. The last question in this section did not apply to any of the respondents.

Section L: Charges

Seven institutions do not charge users to use preserved materials. Four institutions charge a fee for copies; two institutions charge users a fee to use preserved materials when the material is accessed through the web site. Charges are calculated by time (1), by volume of material (2), by intended use (1), (e.g., commercial, academic), students, or as an institutional charge (1). One institution collects charges by credit card, three send invoices, and one is in the process of determining the procedure. One institution collects charges both by credit card and by invoice. Several respondents did not know how the charges were collected in their institutions.

Section M: Reproduction and Copyright

Ten institutions preserve material in copyright, while one institution does not. The question did not apply to two institutions. Of the institutions that preserve material in copyright, eight stated that this was done under legal provisions for their institution, ten with the owner's agreement, and one without formalities.

Six institutions do not own the copyright for the electronic form of the records, while the question did not apply to seven respondents. Ownership of the copyright resides with the creator, owner, or government, depending on the institution.

Users are permitted to download material to a PC (2); download to a local network (LAN) (2); or download to a general network (WAN) (2). One respondent said that the institution does not permit users to download material. The question did not apply to nine respondents.

With respect to electronic management systems used to control copying, one respondent stated that that option was being researched, while four respondents stated that their institutions do not have electronic management systems of that type in use. The question did not apply to eight respondents.

Section N: Preservation Policies

Five respondents do not have a general preservation policy in place that includes electronic records, while three respondents stated that their institutions did have such a policy in place. Two respondents indicated that their institutions had such a policy in development or being researched, while three respondents said that the question did not apply.

Regarding question 15.1, policies for formatting, refreshing, etc., one institution has a policy in place that includes various types of preservation techniques, while four institutions do not. Respondent 10 stated that the institution has no formal policy, but that the established practice is to re-format to standard non-proprietary formats. Two respondents stated that their institutions did not have policies finalized as yet. Additional discussion of preservation policies is provided below.

Definitions, Costs, and Policies

Three questions in particular yielded responses that we believe have ramifications for preservation planning: definitions of preservation (Section A, *Program and Policy*), categories of costs (Section H, *Costs*), and preservation policies (Section N, *Preservation Policies*).

Definitions of Preservation

In the section on *Program and Policy*, we asked: “How do you use the word ‘preservation’ at your institution?” In other words, what definition does your institution associate with the term “preservation”? We included this question because we felt that such definitions might have a bearing on approaches that projects take to developing long-term retention strategies. For example, depending on your view of preservation you might select one approach over another. Since we interviewed archivists and librarians from six countries representing a dozen projects or institutions, we anticipated getting a range of responses. As we continue this study into the next phase of this research, we will try to determine whether or not the definitions of preservation continue to evolve.

Eleven respondents defined “preservation.”¹⁰ From their definitions some key phrases emerged:

- Respondent 1: “Preservation for paper records is a regime which tries to slow entropy and avoid degradation. For digital records, it is to preserve the document to perpetuity. Digital Preservation includes issues of authenticity.”
- Respondent 2: “Preservation means ensuring the object is accessible over the long-term.”
- Respondent 4: “Preservation covers all activities directed towards ensuring the ongoing accessibility to the information content of the records. Hence, we consider the ambient conditions in our repositories as a preservation issue, along with the specifications of the media on which recorded information is stored. Migration of digital objects is thus a preservation strategy.”
- Respondent 5: “The ability to discover, access, and present electronic records through arbitrary changes of technology. We can preserve things forever.”
- Respondent 7: “Forward migration or prospective preservation to whatever new technologies exist. [We are beyond] thinking about ‘x’ number of years of preservation.”
- Respondent 8: “Enabling long-term access to materials.”
- Respondent 9: “Ability to present the record unchanged repeatedly.”
- Respondent 11: “Everything you have to do to guarantee you can deliver records [and] respecting the sanctity of the original order.”
- Respondent 13: “Making collections useful to scholars in the future.”

¹⁰ These have been slightly copyedited.

The respondents' key phrases fall into three components of preservation: preservation processes, length of time for retention, and preservation outcomes. Overall, the responses demonstrate a shift taking place 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, such as proper environmental conditions and storage facilities or proper housing for items, can be taken to insure longevity in paper-based materials. When such measures are put in place, materials can last for a long time, barring disaster, infestation, or careless use. While paper-based materials also require all-the-time care, benign neglect is not always harmful. In fact, in some instances it is better *not* to treat an item. Digital materials, on the other hand, require constant refreshing, reformatting, migrating, etc. These represent much more pro-active and costly endeavors. For digital materials, neglect may result in total loss.

To contextualize the respondents' definitions, we offer published definitions of preservation culled from the archives and library fields, and reflecting a shift that has taken place since the 1970s.

The Paper-Based Perspective

Frank Evans, et al. 1974

A. *The basic responsibility to provide adequate facilities for the protection, care, and maintenance of archives, records, and manuscripts.* B. *Specific measures, individual and collective, undertaken for the repair, maintenance, restoration, or protection of documents.*¹¹

Ratcliffe Report, 1984

Strictly, *all the steps taken to protect materials*, that is including conservation and restoration, but often used in reference to the treatment of materials on first entering the library; *it is preventive rather than remedial.*¹²

Transition to Digital

IFLA, 1986

Includes *all managerial and financial considerations* including storage and accommodation provisions, staffing levels, policies, techniques and methods involved in preserving library and archive materials and the information contained in them.¹³

¹¹ Frank Evans, et al., "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers." *American Archivist* 37 (July 1974): 415–33.

¹² F.W. Ratcliffe with D. Patterson. *Preservation Policies and Conservation in British Libraries: Report of the Cambridge University Library Conservation Project*. Library and Information Research Report, 25 (London: The British Library, 1984).

¹³ J. M. Dureau and D. W. G. Clements, *Principles for the Preservation and Conservation of Library Materials*. IFLA Professional Reports 8 (The Hague: IFLA, 1986).

Paul Conway, *American Archivist*, 1990

Archival preservation is the acquisition, organization, and distribution of resources . . . to ensure adequate protection of historical information of enduring value. . . . Archival preservation encompasses planning and implementing policies, procedures, and processes that together prevent further deterioration or renew the usability of select groups of materials. Archival preservation management, when most effective, requires that planning precede implementation, and that prevention activities have priority over renewal activities.¹⁴

Digital

Society of American Archivists, 1997

Preservation of digital information is not so much about protecting physical objects as about specifying the creation and maintenance of intangible electronic files whose intellectual integrity is their primary characteristic. Preservation in the digital world is not exclusively a matter of longevity of optical disks, magnetic tape, and newer, more fragile storage media. . . . The viability of digitized files is much more dependent on the life expectancy of the access system. . . .¹⁵

International Council on Archives, 1997

An electronic record is preserved if and only if it continues to exist in a form that allows it to be retrieved, and, once retrieved, provides reliable and authentic evidence of the activity which produced the record.¹⁶

Kenney and Rieger, 2000

Digital Preservation means retaining digital image collections in a usable and interpretable form for the long term. While “long-term” suggests an indefinite future, David Bearman interprets it more usefully as ‘retention for a period of continuing value.’¹⁷

Archival and library definitions have shifted from the physical care and protection of materials to retaining them in retrievable form for an indefinite amount of time. In the paper-based information world, librarians and archivists sought to preserve books and documents for five hundred years or more. As is apparent from both the study respondents and the professional literature, professionals now think about maximizing “useful life” or preserving digital

¹⁴ Paul Conway, “Archival Preservation Practice in a Nationwide Context.” *American Archivist* 53 (Spring 1990): 206.

¹⁵ “The Preservation of Digitized Reproductions,” (Chicago: Society of American Archivists, 1997) <<http://www.archivists.org/statements/digitize.asp>> (May 22, 2002).

¹⁶ Committee on Electronic Records, *Guide for Managing Electronic Records From an Archival Perspective* (Paris: International Council on Archives, February 1997), 32 <<http://www.ica.org/eng/mb/com/cer/index.htm>> (May 22, 2002).

¹⁷ Anne R. Kenney and Oya Y. Rieger, *Moving Theory into Practice*, 135.

documents “forever” through emulation or forward migration, but without the emphasis on a specific number of years.

Further analysis of our data indicates that archivists and librarians view preservation through different lenses. This reflects a fundamental difference in the archival and library professions. Librarians tend to be custodians of printed materials that are not unique. Librarianship carries custodial responsibilities; but, with the exception of special collections, missing or damaged items can usually be replaced. Therefore, librarians often view their materials in terms of immediate utility. In the archival arena, when a record is gone, whether because of an accident or a disposition schedule, it is really *gone* and cannot be replaced. Archivists have responsibility for one-of-a-kind records, which are housed in a repository. In current practice, the repository and the object cannot be divorced. This relationship differs from libraries and printed materials. In archives, long-term accessibility to the records may be mandated by legal warrant and business processes, and more broadly, by the need to maintain societal memory. The impact of electronic records may have an effect on the requirements that the repository and the object remain together in archives. In the digital environment, *both* librarians and archivists have responsibility for documents and records that are born digital. These digital assets are susceptible to obsolescence and incompatibility.¹⁸ Therefore, the integrity and authenticity of digital objects are of mutual concern to both professions. As librarians and archivists work closely on long-term retention strategies, the definition of preservation may shift to accommodate both professional perspectives.

Costs

In the section of the questionnaire on costs, we asked, “What do you estimate are the costs to preserve the records?” Responses included staff, equipment, space, energy, and other related costs. In essence, we were asking, “What is it going to cost the institution to preserve, maintain, and provide access to electronic records?” We thought this was an important question because for many institutions and projects, knowing what the bottom line is, is *the* major factor which influences decisionmaking, and determines goals and objectives, as well as the strategies to meet them. Knowing about costs helps repositories lobby with parent institutions or funding agencies. What is the role of costs in situations where, because of legal requirements, archives do not have a choice about what they preserve or even how they preserve records? The majority of the managers we interviewed is gathering financial data now and plan to report

¹⁸ Jeff Rothenberg is now advocating use of the term “inherently digital” to characterize executable, dynamic, and interactive materials. See his *Using Emulation to Preserve Digital Documents* (The Hague: Koninklijke Bibliotheek, July 2000) <www.kb.nl/kb/pr/fonds/emulation/usingemulation.pdf> (May 22, 2002).

costs as part of their projects' results. Only a few projects are far enough along to have developed cost figures. The interviewees ranged from large national archives to projects developing testbeds. The costs for electronic records preservation ranged from \$10,000 to \$2.6 million per year. Cost categories include staff, consultants, facilities, equipment, storage system monitoring, staff access, and research and development.

Most of the projects are currently funded through initial allocations, and some of these figures reflect the impact of early research and development costs, which could also account for the wide range of costs. In fact, as one respondent said, the costs for his project might be reduced by as much as half during the following year. This question will be followed up as part of the second phase of the research interviews. It will be interesting to see what the forecast figures for preservation, storage and staffing actually turn out to be, especially when the initial costs of research and development are reduced over time.

At the time of these interviews, none of the respondents had yet gathered enough information to determine the categories of preservation costs or cost modeling protocols.

Sources of funding include various government agencies—EU (European Union), NSF (National Science Foundation), NPACI (National Partnership for Advanced Computational Infrastructure), NEH (National Endowment for the Humanities), NHPRC (National Historical Publications and Records Commission), NARA (National Archives and Records Administration), and JISC (Joint Information System Committee). As always, the question remains as to what extent the source(s) of funding have shaped the research agenda and from there, the future.

The follow-up study will gather data on the further development of a preservation cost model. So far, cost modeling for digital projects has received scant attention. The present focus appears to be on budgeting for digital conversions rather than preserving authentic electronic records. In addition, there is scant literature in the area of cost models for electronic records that are born digital. Two exceptions are studies by Tony Hendley and by Kelly Russell and Ellis Weinberger. Hendley, in his report on the *Comparison of Methods & Costs of Digital Preservation* provides a "Table of Digital Preservation Cost Elements," which was compiled by Neil Beagrie, Daniel Greenstein, and the Arts and Humanities Data Service.¹⁹ In it, the cost elements involved in developing and preserving digital collections are keyed to the life cycle stages of a digital resource.²⁰ In their study *Cost Elements of Digital Preservation*, Russell and Weinberger posit that the ongoing costs of digital preservation span a more

¹⁹ Tony Hendley, *Comparison of Methods & Costs of Digital Preservation* (London: British Library Research and Innovation Centre, 1998), 96. <<http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>> (May 9, 2002).

²⁰ Hendley, *Comparison of Methods & Costs of Digital Preservation*, 65.

extended timeframe than traditional preservation and will therefore require resource commitments of a different nature. Different strategies may necessitate different costing time frames and schedules. Russell and Weinberger state that current cost models have yet to reflect this more complex environment. They further state that, “The creation of a digital object is the true starting point for digital preservation.”²¹

To estimate a budget for image acquisition, Anne Kenney and Oya Rieger refer to the “RLG Worksheet for Estimating Digital Reformatting Costs” in their book *Moving Theory into Practice: Digital Imaging for Libraries and Archives*.²² The Worksheet, in combination with an assessment of costs derived by Cornell’s Department of Preservation, identified costs for image acquisition in six cost categories. These costs include personnel, equipment, cataloging, supplies, contingency and overhead/indirects.

To facilitate the development of a preservation cost model, a number of categories may be adapted from traditional cost models. These categories might include: costs for providing access to the materials; costs related to long-term creation and maintenance of digital materials, production of metadata, personnel, equipment, cataloging, supplies, contingency (e.g., emergency/unforeseen events), overhead, administration, and research and development.

One respondent provided information about plans to form a consortium of institutions to form a National Preservation Center. This idea should be explored not only because of its potential for cost-effectiveness of preservation, but also for the opportunities to enrich the library, archival and museum professions, which may occur as a result of providing a forum for communication across institutional settings and domains.

In a speech for directors of the Association of Research Libraries, Clifford Lynch stated,

The fundamentally hard things about managing bits into the future mostly aren’t technical; they’re economic and organizational. Bits need care and feeding. They don’t do well with benign neglect. This means that we need to come up with financial models to keep these bits cared for and healthy as they are migrated into the future. We don’t lose a lot of bits to technical failures in a well-managed environment, but we lose a lot due to financial or organizational failures to maintain that well-managed, caring environment on a continual basis.²³

This quote helps to emphasize that technical processes cannot be separated from economic issues. The library and archival professions have not fully grap-

²¹ Kelly Russell and Ellis Weinberger. *Cost Elements of Digital Preservation*. <<http://www.leeds.ac.uk/cedars/documents/CIW01r.html>> (May 9, 2002).

²² Kenney and Rieger, *Moving Theory Into Practice*, 166.

²³ Clifford Lynch, “Strategic Issues: Technology, Trends and Solutions.” In *Preserving Digital Information*, Vantage Point series, pp. 3–4. EBSCO Subscription Services, 2000.

pled with the economic influences on preservation decisions. It is necessary to identify concepts and approaches for evaluating the full economic impact of long-term retention. Institutional, national, and multi-national policies must be put in place to assure preservation in perpetuity.

Preservation Policy

We concluded the survey with the following three issues:

1. Do you have a general preservation policy that includes records in electronic form?
2. If not, do you have a policy for reformatting, refreshing, migrating, emulating, or bundling data to newer technological platforms?
3. Please describe any policies you might have that relate to preservation of electronic records.

Only three of the projects/institutions indicated that they have policies in place; two others are revising existing policies to include electronic records; and one is currently developing a policy that includes multiple media. Two of the research projects indicated that policy development would be an outcome of their research.

During round two of the preservation survey, we will be interviewing key informants/experts who may shed more light on policy issues. We will try to ascertain whether or not international concern about the longevity of digital information is being followed up in policymaking arenas. We suspect that policy is lagging far behind the development of standards, because the development of good public policy requires the appropriate political climate as well as the cooperation of numerous stakeholders. Further, there must be a legal environment that enables the preservation of digital information. Yet laws may vary. For example there are significant differences between the Berne Convention (an international copyright treaty) and United States copyright law. These types of discrepancies may impede the development of consistent, rational public policy.

Conclusion

At present, the interviews indicate three broad themes. First, the perception of what preservation is goes beyond library and archival practice to the media being preserved. Because electronic material is inherently ephemeral, and the timeframe involved to preserve and provide access to this material extends to perpetuity, we expect that traditional definitions of preservation may not apply. Indeed, a shift is already apparent.

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 projects. One respondent, who

is fully funded by his government, put it succinctly when he said, “We haven’t yet been asked to measure costs! We don’t need to justify costs. Fixed costs are unknown.” Another respondent said, “The result will be cost determinations.” And a third answered that costs “should be a result of the current testbed project [that they are engaged in].” We believe that the problems posed by preserving authentic electronic records permanently (or as long as possible), requires the development of a cost model, which will be unique and not a hybrid of existing digital conversion cost models. We agree with Hendley, Russell, and Weinberger that preservation begins at the creation of the electronic material. A cost model for preserving authentic electronic records will need to reflect this perspective, which differs from the traditional preservation point of view. Costs, however, must not single-handedly be used to justify *not* preserving otherwise valuable records.

We found that staff and equipment costs are the most consistent hard figures available so far. Of course, these will vary over time, and will ultimately connect with the development of forecasting strategies. Many of the projects are nascent, and we suspect that for them, answering the survey questions was essentially a theoretical exercise. As the institutions and projects progress, we expect to be able to gather hard data during rounds two and three of the survey. By the conclusion of round three, we expect to have a substantial amount of information about institutions and projects that will have been active for at least three years. From this, we hope to develop a cost model for preserving authentic electronic records, which can be applied to archives and libraries, and perhaps to other communities of practice.

Finally, 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 project, instead of concurrently with its progress? We suspect that meeting the technological challenges of preserving electronic records is more of a priority within these institutions than developing policy and wonder whether, as a result, the overall progress in this new arena will be more uneven than is necessary. Several institutions that responded to our survey have had active programs for a long time, and we note that often policy has just evolved, rather than being strategically planned. It is practically impossible to set policy completely at the outset of a project—especially policy in such a complex area as the preservation of electronic records. Policy will naturally evolve rather rapidly at the outset of a 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 newly conceived. In fact, policy must also drive technological development. When the program is in “full swing,” policy will have reached a point at which it is now well thought out, though still subject to modification, as the program requires.

In the subsequent phases of our survey, 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 role of the stakeholders and the influence of the legal and political environments that provide the context in which policy is formed.

We note that one project we included in round one has discontinued its research because funding ran out. This particular project was unique in that it was exploring the preservation of multimedia material. The gap in potential knowledge that could have been disseminated is a loss. But it is also a reminder that the adoption of any new technology depends on politics, funding, and timing.

Round two of the preservation survey will focus on expanding our knowledge in several areas. These include staffing and personnel—where are future specialists in preserving electronic records going to come from? Another expanded area in the survey will be cost activities. Because some aspects of the programs we studied, such as charges, access, reproduction, and copyright were only in the early stages of development, we were able to gather very little substantive information. As a result, these sections will probably drop out of round two. However, we will re-evaluate their inclusion for round three. We also intend to explore in more depth why certain questions did not apply to some respondents.

No matter which preservation method is chosen, cost will become a factor in making a management decision regarding preservation of electronic records. We have also expanded the area of our survey that asks for information on description/documentation of preservation processes, as well as the section on preservation policy. In addition, we have developed a second survey instrument that we will use to interview key informants/experts, whom we define as “individuals who provide useful insights in the fields of preservation and/or archives.” These experts may work in a variety of settings including—but not limited to—universities, government archives, foundations/granting agencies, industry, professional organizations, or think tanks, or who serve as consultants. Most have extensive national and international experience that enables them to provide the long view of preservation as well as placing them at the forefront of their professions.

* * * * *

We began our study by asking four research questions:

1. What methods, procedures, and rules of long-term preservation are in use or being developed?
 - a. Which of these meet the conceptual requirements for authenticity?
 - b. Which methods of long-term preservation need to be developed?
 - c. Which of these methods are required or subject to standards, regulations, and guidelines in specific industry or institutional settings?

2. What is the meaning of *preservation*?
 - a. Does the meaning change when it is applied to electronic rather than paper-based records?
3. Will current strategies for preserving electronic records ensure longevity and authenticity?
4. How are costs for the preservation of electronic records derived? Have effective cost models been developed?

Our survey provided us with only partial answers to Questions 1 and 3. We identified a number of preservation techniques that are currently in use—including migration, emulation, and robotics—but not one of these techniques could be considered to meet the conceptual requirements for authenticity. Until these methods are further developed and standardized, they cannot be relied upon to ensure the long-term preservation of electronic records. As for question 1.b, “Which methods of long-term preservation need to be developed?” the answer is that they all do. It is still too early in the development of all these techniques to fully evaluate them. Regarding question 1.c, “Which of these methods are required or subject to standards, regulations, and guidelines in specific industry or institutional settings?” there is no simple answer. The projects represented in our survey are developing standards and guidelines. Some of the institutions we interviewed are waiting to see the results of these projects before committing to a particular strategy. We hope to be able to answer question 1 more fully in subsequent rounds of this research.

Question 2 yielded richer results. It is clear that professionals are revising their definitions of preservation from a once-and-forever approach for paper-based materials to an all-the-time approach for digital materials. Preservation must now accommodate both media and access systems. Finally, while we once tended to think about preserving materials for a particular period of time—for example, permanent/durable paper was expected to last for five hundred years—we now think about retaining digital media for a period of continuing value.

Meaningful answers to question four regarding costs for the preservation of electronic records, must also wait until rounds two and three. Our survey revealed that in the rush to develop the technological processes necessary to preserve authentic electronic records, cost issues have often been pushed aside. This is in part because ample government and foundation funding is allowing some institutions to defer cost modeling. Many respondents reported that they are beginning to study the cost implications, and we hope to gather more information in the next round.

As a result of the information we will gather over the next two years about evolving preservation practices, we expect to strengthen the foundation underlying the development of the preservation function model, particularly those aspects which concern preservation, storage, and access to authentic electronic records over time. We also hope to provide insights which will contextualize the

work of projects and institutions around the world, and which will ultimately provide a pool of knowledge that will benefit us all.

* * * * *

When John Steinbeck completed *Of Mice and Men*, he described it to his publisher as an experiment, adding, “don’t publish it if you don’t like it.” So unsure was the author of his work that he did not even want to read the proofs. Christopher Morley, describing for *Book-of-the-Month Club News* the publication of Steinbeck’s book, wrote that “in just such casual ways, in this our world of obliquity and squint, do masterpieces happen.”²⁴ “Obliquity and squint” captures the notion of looking at something without fully understanding or seeing it. It is as apt a description of electronic preservation as it is a description of the chance publication of an enduring work of literature. In this period of incunabular electronic information, it is difficult to understand all the potential or all of the pitfalls of the newest forms of our cultural heritage. Until we do, however, electronic preservation itself will be seen as oblique.

²⁴ Quoted in Al Silverman, ed., *The Book-of-the-Month: Sixty Years of Books in American Life* (Boston: Little, Brown, 1986), 49.

Appendix I: Definitions

Each definition is followed by a parenthetical source reference.

Preservation: the processes and activities that stabilize and protect objects so that they will be permanent and durable, or as long lasting as it is possible to make them (authors).

Digital Preservation: the processes and activities which stabilize and protect reformatted and “born digital” authentic electronic materials in forms which are retrievable, readable, and usable over time (authors).

Digital Preservation Techniques

Refreshing: periodically moving records from one storage medium to another. It is a preventive measure and, because of rapid media obsolescence, it will be a necessary strategy for some years to come (authors).

Migration: the process of moving records from one hardware and/or software platform to another (authors).

Collection-Based Persistent Object Preservation (a.k.a., **Knowledge-Based Persistent Object Preservation**): “. . . the retrieval or instantiation of the collection onto new technology. . . . [t]he processes used to ingest a collection, transform it into an infrastructure independent form, and store the collection in an archive comprise the persistent storage steps of a persistent archive. The processes used to recreate the collection on new technology, optimize the database, and recreate the user interface comprise the retrieval steps of a persistent archive. The two phases form a cycle that can be used for migrating data collections onto new infrastructure as technology evolves” (Reagan Moore, et al., “Collection-Based Persistent Digital Archives—Part 2,” *D-Lib Magazine* 6 (April 2000): 1 <<http://www.dlib.org/dlib/april00/moore/04moore-pt2.html>> (May 22, 2002).

Emulation: an applications software approach that recreates the technical environment required to view earlier programs. Such software can theoretically mimic every type of application ever written and be run on current computers (authors). See also, Stewart Granger, “Emulation as a Digital Preservation Strategy,” *D-Lib Magazine* 6 (October 2000): 1–12 <<http://www.dlib.org/dlib/october00/granger/10granger.html>> (May 22, 2002).

Technology Preservation: “preserving the technical environment that runs the system, including software and hardware such as operating systems, original application software, media drives, etc. While technology preservation means preserving the technical environment rather than reengineering it, as emulation does, many of the same issues apply to both” (Kenney/Rieger, *Moving Theory into Practice*).

Bundling: “. . . taking objects such as Word documents and, by using software, creating bundles of documents on an independent platform. This is another form of emulation” (Public Records Office interview).

Universal Preservation Format (UPF): “Deals with multiple formats. It takes a snapshot of the information and gives you everything you need to view it. It is difficult to capture the whole content, e.g., interactivity, without the original equipment. Closer to emulation than migration” (UPF is a technique developed at WGBH in Boston; WGBH interview).

Digital Archaeology: “accessing digital materials where the media has become damaged (through disaster or age) or where the hardware or software is either no longer available or unknown.” Seamus Ross and Ann Gow refer to this as “*post hoc* rescue” in their *Digital Archaeology: Rescuing Neglected and Damaged Data Resources* (Glasgow: Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, February 1999): iv <www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2con.pdf> (May 22, 2002).

Robotics: the use of robots to download electronic documents. The process of downloading in and of itself does not preserve the records (authors).

Archival Terms

Record: a document made or received and set aside in the course of a practical activity (InterPARES Glossary, <<http://www.interpares.org/documents/InterPARES%20Glossary%202002-1.pdf>> (May 22, 2002).

Authentic record: a record that is what it purports to be and is free from tampering or corruption (InterPARES Glossary, <<http://www.interpares.org/documents/InterPARES%20Glossary%202002-1.pdf>> (May 22, 2002).

Digital record: a record that now exists in electronic form though it may or may not have been created in electronic form. For example, a *digital record* may have been created on paper and digitized later. Subsequent digitization may remove or deplete its “recordness” (authors, Gilliland-Swetland, *Enduring Paradigm*).

Electronic record: A record that is created (made or received and set aside) in electronic form (InterPARES Glossary, <<http://www.interpares.org/documents/InterPARES%20Glossary%202002-1.pdf>> (May 22, 2002).

Other Terms

Program: an ongoing set of services, around a common goal or activity, usually located within a single institution (authors).

Project: a specific undertaking or research endeavor, usually with special funding. Projects may take place within single, institutional programs, or at more than one site. Projects usually take place within a finite period (authors).

**Appendix 2: InterPARES Preservation Survey List
of Sites—Round I**

1. Public Records Office, England (A)
2. Cedars Project (P)
3. Ministry of the Interior, Netherlands (A)
4. National Archives of Australia (A)
5. San Diego Supercomputer Center (P)
6. WGBH – UPF (P)
7. Large American University Library (L)
8. Creative Archiving at Michigan and Leeds: Emulating the Old on the New (CAMiLEON) (P)
9. National Archives of a northern European country (A)
10. National Archives of Canada (A)
11. National Archives and Records Administration (NARA) (A)
12. Cornell University (P)
13. Internet Archive (P)

A = Archives (6)

L = Library (1)

P = Project/Program (6)

**Appendix 3: Questionnaire and Tabulated Responses—
Round I**

Questionnaire on Preservation Strategies for Electronic Records

A. About the Archive/Institution(s)

1. Name of Institution or Project: n = 13
 archives = 6
 projects/programs = 6
 library = 1
2. Person (s) responsible for the program (Name (s) and Title (s)):
 n = 13
 2.1. Address: n = 13
 Countries represented: England (3), Netherlands (1),
 Australia (1), US (6), Sweden (1), Canada (1)

B. Program and Policy

3. Please describe your institution's program or activities related to preserving digital objects over long periods of time.

- 3.1. When did your institution's program or activities begin? n = 13, n/a = 2,
 Prior to 1970 = none
 1970–1979 = 4
 1980–1989 = 1
 1990–1999 = 5

- 3.2. Describe any institutional issues which impact upon the program.

- 3.2.1. Describe in broad terms what methods or techniques you are exploring or using for digital preservation. n = 13, n/a = 2
 Migration = 4
 Emulation = 2
 Knowledge-Based Persistent Object Preservation = 3
 Bundling = 1
 Refreshing = 1
 Digital Archeology = 1
 Preservation Copying = 1

Physical Preservation = 2

Robotics = 1

- 3.3. Describe the digital materials your institution is preserving. n = 13, n/a = 1, none = 1 (funding ran out) Examples of digital materials include:

All “born-electronic” records, e-mails, transcripts, evidence, annotated documents. Paper records are scanned in.

Passive storage of data files (650,000).

Copyright deposits of print materials, i.e., digitized materials, image collections.

As wide a range of materials i.e, research library items, as possible and is looking at CDs.

Are working with an art museum consortium—The art and information about the object, including 2 levels of metadata:

(1) on the object—the digital record, and (2) on the digitized object.

Records of every federal government agency, presidency and Congress.

Serials, books, image files.

Chuckie-Egg games; spreadsheets; databases; other data formats.

Statistical datasets, tax records, other government records the archive is required by law to take.

Geomatics : geographical information in automated systems, climate information

Databases containing text.

- 3.3.1. Do you consider any of these materials to be records?

n = 13, n/a = 6

Yes = 6

No = 1

- 3.4. Do you make any special provisions for preserving records, as opposed to other types of digital materials? If so, what? n = 13, n/a = 7

Yes = 1

No = 5

- 3.5. Has the program or activity reached the point of either testing or evaluating any of the methods or techniques you are using? If so, what are the results to date?

- 3.6. Have you identified any problems, difficulties or threats to the

integrity of the digital materials resulting from the use of these methods or techniques? If so, please describe.

3.7. How do you use the word “preservation” at your institution? In other words, what definition does your institution associate with the term “preservation”?

C. Specifics of Preservation Technique/Method or Strategy

- 4. What preservation technique does your program use? n = 13, n/a = 4
- None = 3
- Migration = 5
- Emulation = 1
- Scanning = 1
- Reformatting = 2
- Refreshing = 1
- Microfilming = 1

4.1. How was this method selected?

4.1.1. Is it a hybrid, e.g. a combination of two or more preservation techniques/methods, such as microfilming and scanning? YES NO

- n = 13, n/a = 6
- Yes = 4
- No = 2
- Maybe = 1

4.2. If you are using a hybrid model, how did it evolve?

4.3. From other methods you have tried before or you are aware of that other repositories are using, how is this method different from other methods?

- 4.4. In selecting the preservation method or strategy, have you considered what its effect might be upon the intellectual integrity (e.g., authenticity and reliability) of the digital material?
 YES NO
n = 13, n/a = 4
Yes = 9
No = \emptyset
- 4.4.1. If yes, are you able to prove/demonstrate that the intellectual integrity of the digital material has not been compromised through the preservation process? Please explain.
n = 9, n/a = 5
Yes = \emptyset
No = 4
- 4.5. Is there evaluative data on the efficacy of this preservation method/model? Please describe. n = 13, n/a = 6
Yes = 3 ((1) no comment; (2) research report—fits OAIS model; (3) Audit trails are kept. They are looking at moving up to a document management program (ERM) to support audit trail development)
No = 3

D. Selection for Preservation

5. Which of these criteria guide selection of materials for preservation? (Check all relevant) n = 13, n/a = 1
 Historical/cultural value (9) Save space (3)
 Legal requirement to preserve (6) Research into preservation processes (5)
 Retard deterioration (5) Commercial use (\emptyset)
 Increase access (6) Other reasons (please specify) (4)
(1) Materials publishers wanted to let them have; (2) Institutional requirements; Asset management considerations; (3) To support the curriculum; (4) Sampling.
- 5.1. Where did the materials you selected come from? n = 13, n/a = 1
 Parent institution (5) Government agencies (6)
 Other (specify) (5)
(private, commercial publishers, politician's private papers, private individuals, commercial entities)
 Collaborating institutions (3) Other organizations or associations (2) (corporations (private sector))

E. Cooperation

6. Did you cooperate with other organizations to develop your program?
 YES NO

n = 13

Yes = 12

No = 1

6.1. If so, which? (Check all relevant) n = 13, n/a = 1

- Archives (9)
- Libraries (9)
- Public companies (5)
- Museums (3)

Other: government agencies, other programs and projects, for profit and non-profit educational institutions; and universities.

6.2. Is your cooperation National (10) International (11)

- Local (2)
- Shared facilities (1)
- By institutional type (Ø) [churches, labor unions, etc.]

n = 13, n/a = 1

6.3. How is the work distributed? n = 13, n/a = 2

- Equally (5)
- Work distributed in a different way? (6)

6.3.1. Please describe.

6.4. If your program is collaborative, how did it evolve?

6.4.1. Please describe the strengths and weaknesses of the collaboration.

F. Staffing

7. Who is involved with the program and in what capacity (ies)?

7.1. Describe their duties.

G. Technical Questions

8. Is preservation carried out by the institution (in-house) (11)
 commercial vendor/contractor (2)

n = 13, n/a = 2

Note: One respondent uses both.

8.1. Describe any pre-preservation preparation of records.

8.2. What do you consider to be the strong points of your institution's preservation methods or techniques?

8.2.1. What do you consider to be the weak points of your institution's preservation methods or techniques?

8.3. What quality control methods are applied to the preservation process or activity?

8.4. How are you storing the electronic records that have been preserved?

H. Costs

9. What do you estimate are the costs to preserve the records? (Please include staff, equipment, space, energy and other related costs)

9.1. If applicable, please describe the categories of your preservation costs.

9.2. What are the sources of funding for the program, and how are they allocated?

I. Preserving Records (if applicable)

10. How are preserved records organized?

10.1. How is provenance respected?

10.2. Are there any restrictions on access to the records? If so, how are they enforced?

J. Description/Documentation of Preservation Processes

11. Describe record-keeping for the preserved material.

11.1. Are preserved materials described according to a recognized standard? YES NO

n = 13, n/a = 4

Yes = 8

No = 1

11.1.1. If yes, which one? n = 8, n/a = 5

ISOD-G, various, EAD, MARC, modified LCSH, RAD,
Dublin Core

11.2. Are the records for preserved materials and the original
 the same independent of each other

11.3. How is metadata used to describe preserved materials?

K. Access to Preserved Records

12. Are the preserved records available only on-site (3) only within the institution (1) through a website (6) other (specify) (2) ((1) hard copy; (2) copyright protected material available to the [academic] community only.)

n = 13, n/a = 4

12.1. If available through a website, please give the URL

12.2. Is the archival workstation equipped with access control mechanisms billing software

n-13, n/a = 12

Dedicated machines only: 5 workstations, some online, some connected to secure servers.

12.3. Is the archival workstation connected to internal institutional servers an Intranet the Internet

n = 13, n/a = 13

L. Charges

13. Do users have to pay to use the preserved material? YES
 NO

n = 13, n/a = 4

Yes = 2

No = 7

Fees are charged for copies. (4)

13.1. If yes, on-site outside the Institution (1) when accessed through the website (2) other (specify)

n = 13, n/a = 10

If they create remote access to electronic surrogate documents, they will charge for downloading/browsing, which are value-added services. The services are free onsite.

13.2. If charges are made, how are these calculated?

single charge (1) by time (1) by volume of material (2) by intended use (1) (commercial/academic/students) customized service (please describe) other (give details) (1) (institutional charge)

n = 13, n/a = 10

13.3. If charges are made, how are they collected? invoice (3) cash at point of use credit card (1) electronic accounting other (give details)

n = 13, n/a = 9, TBD = 1

M. Reproduction and Copyright

14. Do you preserve material in copyright? YES NO

n = 13, n/a = 2

Yes = 10

No = 1

14.1. If yes, is this done under legal provisions for your institution (8) with the owner's agreement (1) by paying the owner a fee under license without formalities (1)

n = 13, n/a = 3

14.2. Does the institution own the copyright for the electronic form of the records? YES NO

n = 13, n/a = 7

Yes = Ø

No = 6

14.2.1. If no, who does? _____

14.3. Are users allowed to do any of the following?

Ø make printouts (3) Ø download to a PC (2) Ø download to a local network (LAN) (2) Ø download to a general network (WAN) (2)

n = 13, n/a = 9

No = 1

14.4. Are any electronic management systems used to control copying? Ø YES Ø NO

n = 13, n/a = 8

Yes = 1 (Research & Development TBD)

No = 4

14.4.1. If yes, which ones?

N. Preservation Policies

15. Do you have a general preservation policy that includes records in electronic form? Ø YES Ø NO

n = 13, n/a = 3

Yes = 3

No = 5

In development/revision = 2

15.1. If not, do you have a policy for reformatting, refreshing, migrating, emulating, or bundling data to newer technological platforms? Ø YES Ø NO

n = 13, n/a = 6

Yes = 1

No = 4

Policy not finalized = 2

15.2. Please describe any policies you might have that relate to preservation of electronic records.

Do you have any supporting documentation that you can share with us? e.g., policies, specifications, non-proprietary information? Ø YES Ø NO