InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: **General Study 05 Final Report: Round 2: The Preservation of Authentic Electronic Records – *Ad hoc,* Inconsistent, or Strategic?**

Status: Final (public)

Version: 1.0

Submission Date: November 2003 (Revised February 2004)

Release Date: September 2007

Author: The InterPARES 2 Project

Writer(s): Michèle V. Cloonan, Dean

Graduate School of Library and Information Science, Simmons College, USA

Shelby Sanett, Director of Special projects
Amigos Library Services, USA

Project Unit:

URL: http://www.interpares.org/display_file.cfm?doc=ip2_gs05_r2_final_report.pdf

Table of Contents

Abstract

The authors are conducting a three-part study to evaluate current trends in the preservation of digital assets, with an emphasis on electronic records. The study emanated from the authors' work on the Preservation Task Force of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project. This article incorporates the findings of the survey and individual key informant interviews that they conducted from August 2001 through February 2003, as round 2 of the study. Round 2 builds on the 2000-01 round 1 survey that sought to identify and describe strategies for preserving electronic records. In this round the authors found that progress has been made in some areas, while still lagging in others.[1] The full study consists of three phases: round 1 identified and surveyed thirteen institutions, projects, and programs in North America, Australia, and Europe. Round 2 surveyed eight of the thirteen institutions again to follow up on their progress and interviewed eighteen key informants. In round 3 the authors are conducting two case studies drawn from the survey participants in rounds 1 and 2. By the end of the three rounds, the authors will have studied a continuum of activities over a six-year period that constitutes a range of digital preservation strategies. The study will have charted the change in technological developments over this six-year period—developments that have occurred in our survey institutions to meet the requirements of their mandates to preserve digital assets for as long as they are needed.

Prologue

Preserving authentic records in electronic systems is the Holy Grail for archivists. We are compelled to find solutions for problems that we cannot yet fully grasp. Not only is the Grail itself elusive, but the path to it itself is still largely untrodden. The first challenge is identifying what *is* a record, the second is appraising it, and the third is accessioning and preserving it. Other roadblocks include: technological obsolescence, storage media fragility, the manipulability of electronic records, costs, and intellectual property rights (IPR) issues. Then there is the Grail itself: the records which continue to change forms. In a short time we have moved from managing mostly flat files to managing multi-media, dynamic, experiential, and interactive records. What might documents be like in the future? And how will we access and use them?

We continue this quest, of course, because we have a responsibility and a mandate to preserve records for as long as they may be needed. Absent a modern-day Galahad, we need to pursue a variety of strategies. The underlying question remains, how do we preserve digital content that is reliable, authentic, and accessible over time?[2] The purpose of this study is to identify approaches to and ideas about digital preservation. In this report we discuss the findings of part 2 of our study, in which we continue to document extant practices as well as expert opinions on the digital preservation dilemma.

[1] Round 1 and 2 were prepared, in part, with support from the National Historical Publications and Records Commission (NHPRC), the National Science Foundation (NSF), the Social Sciences and Humanities Research Council of Canada (SSHRC), Amigos Library Services, Simmons College, and the State University of New York at Albany.

[2] See Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Chicago: Society of American Archivists, 2004) for the variety of definitions for these terms.

Definitions of Key Terms

This report faces the problem of using a highly specialized vocabulary that is complicated by the fact that our research deals with phenomena from two distinct but overlapping fields: libraries and archives. Practitioners and scholars in these fields deal with similar or identical sources, but with slightly different perspectives. For example, academic librarians tend think about preserving information permanently, while archivists preserve documents and records for as long as they are needed based upon retention schedules. Also, while some archivists might say that "part of a record without regard to its evidential value is not worth keeping at all,"[3] many librarians would say that keeping part of a record—especially if that is all that is extant—is better than keeping no part of it at all. It follows that archival and library vocabularies are not always congruent. Some of the key terms, then, must be defined here, with these differences in mind. It is noteworthy that the terms are obviously similar in meaning, arising from essentially the same source materials. Not only are there subtle but important distinctions among them, but also, archivists and librarians do not always use them the same way, as we have suggested. (The definitions below are taken from the InterPARES 2 *Glossary*.[4])

A **record** is a document made or received and set aside in the course of a practical activity. Thus a record is not merely private information. As one of our key informants, Sue McKemmish, put it, (see *Findings: The Key Informants Interviews*, below), a record has a "fixed content" which can be "re-presented in the structure or form in which it was born." The preservation of authentic electronic records requires that the recordkeeping processes must also be maintained.

An **electronic record** is a record that is created (made or received and set aside) in electronic form.

An **authentic record** is a record that is what it purports to be and that is free from tampering or corruption.

A **digital record** is one that now exists in electronic form though it may or may not have been created in electronic form. A digital record may have been created on paper and digitized later. Subsequent digitization may remove or deplete its "recordness."

To these terms, we add:

A **digital component** is a stored digital object which is necessary to reproduce an electronic record (or other digital asset).[5]

Digital assets and digital content refer to all types of information, text, graphic, image, and multimedia (authors' definition).

[3] Anne Gilliland, e-mail to the authors, January 20, 2004.
[4] Available at http://www.interpares.org/ip2/ip2_terminology_db.cfm.
[5] Thanks to Ken Thibodeau for providing us with this, his working definition. (e-mail to Shelby Sanett, April 22, 2004).

Background

This three-part study grew out of our work on the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 1 Project, a three-year international and multidisciplinary collaborate research effort that examined the preservation of the authenticity of records that needed to be preserved.[6]

As members of the Preservation Task Force, we[7] were asked to document current strategies in use for preserving authentic electronic records. Our study aimed to inform the primary research objective of the Task Force: to develop a model of the process of preserving electronic records.[8] In consultation with the other task force members, we designed a survey instrument and sent it to fifteen institutions or programs (of which thirteen responded). The survey questions were based on the research questions that were the focus of the InterPARES 1 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 these questions, we posed additional ones having to do with definitions, preservation techniques (e.g., refreshing, emulation, migration, robotics), and costs. See Appendix 6 for the list of changes made from round 1 to round 2 of the "Preservation Strategies for Electronic Records" questionnaire.

The findings of round 1 of our study were reported in an article in *The American Archivist*.[9] In that study we found that a number of preservation techniques were in use but that none of them could be considered meeting archival requirements for authenticity. The study also revealed that while developing technological processes to preserve authentic electronic records, almost every institution had deferred costing digital preservation processes and implementing digital preservation policies. This study shows that progress has been made in some areas since 2000, while still lagging in others. For example, institutions are beginning to think about cost issues and models, but have been slow to develop digital preservation policies and plans. As one of our survey respondents in round 2 observed about his own institution, "As long as there is no plan, the risk will be that preservation will be *ad hoc*, inconsistent, and not imbedded in the organization."[10] Hence, the subtitle of this article.

---

[6] For all the findings from InterPARES 1, see the InterPARES Project Web site, http://www.InterPARES.org. (InterPARES 2 (IP 2) is a five-year continuation project that is "applying a multi-method approach to the development of concepts, processes and tools that will help in the securing of a protected and lasting environment for the digital records produced in interactive, dynamic and experiential systems in the course of artistic, scientific, and government activities." Round 2 overlapped with the beginning of IP 2. See also: http://www.InterPARES.org for further information.

[7] The first person plural pronouns in this report refer to its two authors.

[8] For the preservation model, see the final report of the Preservation Task Force, *Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* at: http://www.interpares.org/book/index.htm.

[9] Michèle V. Cloonan and Shelby Sanett, "Preservation Strategies for Electronic Records: Where We Are Now—Obliquity and Squint?" *The American Archivist* 65 (Spring/Summer 2002): 70-106.

[10] Institutional interview #3, conducted September 1, 2001, in Washington, D.C.

In round 2, we used the same categories of questions, though we refined and/or expanded them based on our findings in round 1 (see Appendix 1 for the round 2 survey). We also interviewed eighteen key informants.[11] As we detail in the Research Methodology section below, they had diverse backgrounds and worked in a variety of settings such as universities, government archives, foundations, and professional organizations. Some of them are also researchers and consultants. The informants provided the study with a broader preservation perspective than we could glean solely from the survey respondents, who represented their individual institutions. The questions we posed to the key informants were also more open-ended so that we could elicit as much information as the interviewees wanted to share (see Appendix 2 for the round 2 Key Informant Questionnaire).

Rationale/Purpose of Study

This study continued the research that we undertook in round 1 of the study to assess current trends in digital preservation. Since the results in round 1 revealed that the field is still fluid, we decided not only to continue to monitor the programs (Appendix 3) that we surveyed in round 1, but to place the activities of the surveyed institutions and programs into a broader preservation context. One technique for doing so is to interview experts in a field. We identified 19 key informants (Appendix 4), and were able to interview 18 of them (see below). The institutional interviews that followed the surveys focused on activities within the given institutions. What we hoped to gain from interviewing key informants was a comprehensive perspective from individuals who have many years of experience and expertise. We wanted a perspective that was broader than that of most preservation managers. Therefore we sought to interview people who address preservation issues as part of their overall professional responsibilities. For example, some of our informants fund preservation activities, others conduct preservation research, and still others teach preservation. Some of the interviewees have done more than one of these activities. Interviewing key informants enables us to feel confident that the results of the surveys can be placed within the context of the preservation field overall, because we can compare the results of the survey to the perspectives voiced by our key informants.

The only study that is similar to this one is Beagrie's, *National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands and the United Kingdom* .[12] carried out between our round 1 and the completion of this round. His report was commissioned by the Council on Library and Information Resources and the Library of Congress as part of the latter institution's planning for its National Digital Information Infrastructure and Preservation Program (NDIIPP). Therefore, Beagrie's emphasis is on national digital preservation initiatives primarily in libraries rather than in archives.

[11] We drew the definition of key informants from two sources: John W. Creswell, *Qualitative Inquiry and Research Design: Choosing Among Five Traditions,*1st ed. (Thousand Oaks, CA: Sage, 1998); and, *The American Heritage Dictionary*, 3rd ed. Creswell defines *key informants* as "individuals who provide useful insights into the group and can steer the researcher to information and contacts," 60. *The American Heritage Dictionary* defines *informant* as "One who furnishes linguistic or cultural information to a researcher," 927.

[12] Neil Beagrie, *National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity.* (Washington, D.C.: Council on Library and Information Resources and the Library of Congress, April 2003).

Research Methodology

Our three-part study uses a mixed methods approach: survey and follow-up interviews (round 1); survey, follow-up interviews, and key informant interviews (round 2); and case studies (round 3). Many researchers in the social sciences use mixed methods techniques when it is appropriate to link quantitative and qualitative methods. Creswell describes mixed methods this way:

> . . . [the] approach is one in which the researcher tends to base knowledge claims on pragmatic grounds (e.g., consequence-oriented, problem-centered, and pluralistic). It employs strategies of inquiry that involve collecting data either simultaneously or sequentially to best understand research problems. The data collection also involves gathering both numeric information (e.g., on instruments) as well as text information (e.g., on interviews) so that the final database represents both quantitative and qualitative information.[13]

Other terms for this approach include integrative, synthesis, and multimethod, but "mixed methods" is gaining currency, as is indicated by the list that Creswell provides of current scholars who are using—and have written about—it.

The mixed methods approach best suited our overall strategy to evaluate current trends in and perspectives on the preservation of digital assets. We began the research by gathering data about digital programs and practices (round 1), then repeated the survey with a subset of the first survey participants (round 2). At the same time as we administered the survey in round 2, we developed a detailed view of digital preservation issues through our interviews with key informants. In the process, we collected both closed-ended quantitative data and open-ended qualitative data.

In this round we used a concurrent data collection strategy: the surveys were administered while the key informant interviews were taking place. At the outset, we gave equal weight to the two strategies but, in the final analysis, the study is more qualitative, both because of the small survey size, and the fact that we conducted follow-up interviews with the survey participants. When added to the key informant interviews, most of our data ended up being qualitative. For a study of this nature, which looks at phenomena, policies, and practices in the real world, it seems that the qualitative data we gathered would have more value to the users of this report than would mere numbers.

For the survey portion of this study, we followed the same protocols as in round 1: we sent the surveys to the appropriate contact at each institution, and each person returned the completed survey either electronically or via snail mail. We noted in the responses areas where we needed clarification or amplification. We then interviewed each respondent by telephone or in person. We taped the interviews and graduate students from the School of Information Science and

[13] John W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 2nd ed. (Thousand Oaks, CA, 2003): 19-20.

Policy at the State University of New York (SUNY),[14] Albany transcribed the tapes. We made preliminary corrections and then sent the transcriptions to the respondents for further corrections.

As in round 1, we do not attribute quotes from the survey to specific individuals in the institutions where they work. The institutions are identified in Appendix 3. Quotes from the key informant interviews are used with permission of the interviewees.

Our informants work in different types of institutional settings. Most have had national and international experience, providing a broad view of preservation. Other characteristics of the informants are worth noting:

- Seventeen of the 18 have a minimum of 15 years of experience in their profession, and have experienced the transition from analog to analog plus digital preservation
- One has 10+ years in the profession
- We selected people with wide-ranging perspectives:
    - Academies (faculty and/or researchers)
    - Consultants/Researchers
    - Funders from the private or public sector
    - Those working in government agencies
    - Individuals with preservation experience directly in the field or else as educators or as funders to, or collaborators with, practitioners of the field
    - Representatives from other organizations that are consortial or that sponsor research
- Some of the informants we thought would provide thoughtful answers about the field based on their experience.

The key informants received the questions in advance of their interviews (Appendix 2), but we did not ask them to prepare answers because the questions were designed to be points of departure that would generate discussion. (As it turned out, some of the informants brought notes to the interviews, but this preparation seemed to inform rather than shape the discussions.) Of the eighteen interviews, fourteen were face to face, and four[15] were on the telephone. All of the interviews were taped except one, in which the tape recorder malfunctioned. As with the survey interviews, the tapes were transcribed by graduate students, and then corrected by the interviewers and interviewees.

We interviewed the key informants individually as well as in groups of two or three. We selected participants for the groups who had a variety of working relationships. We tried to select group members whose experience would be complementary. In one interview, the informants worked for the same organization but in different parts of the country; another group was working on a joint research project; another served on a task force together. Still another group's participants worked together in professional associations.

[14] The transcribers were Kevin Glick (now at Yale University), Reg White, and Mark Wolfe. Jonathan Lill, then a student at the Graduate School of Library and Information Science at Simmons College (and now at Columbia University), helped with the editing, and Francesca Marini, a doctoral candidate in the Department of Information Studies at UCLA, assisted with the transcripts and citations. We acknowledge their work with many thanks.

[15] In the last group interview, the interviewer and one interviewee were in New York City, while the other, in California, participated by telephone.

Shelby Sanett alone, Michèle Cloonan alone, or Sanett and Cloonan together conducted the survey and key informant interviews. The surveys and follow-up interviews to the surveys took place between September 2001 and February 2003. The key informant interviews took place between August 2001 and February 2003.

The survey was pre-tested in round 1 and round 2 by members who had served on the InterPARES Preservation Task Force. Since we designed the key informant interview instrument to complement and amplify the survey questionnaire and to have open-ended questions, we did not pre-test it.

Between the surveys and the key informant interviews, we spoke with twenty-six people from Australia, Canada, the United States, and Europe. Our aim was not specifically to conduct a comprehensive international study, but we wanted a broad perspective, and we got it by questioning people beyond the United States. Of course archival and library models are different from country to country; our view is necessarily America-centric.

We designed the survey questionnaire to track specific institutional practices, whereas we asked the key informants more general questions about preservation. There were three overlapping areas: staffing, cost, and policy (see Figure 1), though even in these areas, the emphases were intentionally different. For example, policy in the survey referred to institutional practices, whereas we asked the key informants to address national preservation policies (see Appendices 1 and 2). We obtained far more information than we could use in this study, so we concentrated on the themes that related most directly to the questions.

During the key informant interviews, all of the interviewees cited books and articles. Some of the readings were ones that interviewees found influential, while others were just mentioned as items that the interviewee had read recently but had relevance to the discussion. In some cases the interviewees wrote, or helped to write, the items cited. Since we did not specifically ask the interviewees about their reading, these spontaneous citations are of some interest. Who reads whom? Who cites whom? The answers to these questions appear in Appendix 7, where we list all of the writings that were cited, who cited them, and the frequency with which each item was cited. Although outside the scope of this study, another researcher may wish to explore social networks within the archives or preservation communities.

By using different methods and sources in this study—surveys, key informant interviews, and a literature review—we sought to corroborate our findings for the themes we examine. By triangulating our data, we sought to cross-validate our small sample of institutions and individuals.[16]

[16] See Creswell, page 202, where he defines triangulation as using multiple sources to provide corroborating evidence. In his second edition (2003), Creswell names our approach the "concurrent triangulation strategy" (p. 217). The definition is essentially the same: "two different methods in an attempt to confirm, cross-validate, or corroborate findings within a single study."

Figure 1. Themes Addressed in both the Surveys and Key Informant Interviews by the Questions

Survey Questionnaire	Key Informant Questionnaire
Cooperation	**Cooperation**
Staffing	**Staffing, Education, Training**
Cost (of program and of preservation)	**Costs (of preservation)**
Policies (institutional preservation policies)	**Policy (National preservation policy)**

Boundaries and Limitations

For the survey portion of this research, there were only eight survey participants and eighteen interviewees; this is too small a sample size to be generalizable, however we are able to identify trends or potential trends. As we have noted, though the small number of respondents would yield data insufficient to be meaningful statistically, they did yield a good deal of excellent qualitative information. The small number of participants is also a by-product of the small number of practitioners in the world who can offer significant data for this study.

The participants represented only Europe, North America, and Australia—lack of resources caused us to limit the survey and interviews; therefore we could not include Asia, the Middle East, and South America where we are aware there are digital initiatives in some of those places. Also, we interviewed only participants who spoke English.

Those who use qualitative methods have examined the relationship between researcher and respondent. McCracken writes, "This intimate acquaintance with one's own culture can create as much blindness as insight. It can prevent the observer from seeing cultural assumptions and practices."[17] In this study, familiarity with the field probably helped lead the researchers to some of the interviewees. On the other hand, this same familiarity may have prevented the researchers from asking more questions in cases where they should have.

One possible limitation here is that the researchers have a background stronger in library practice than in archival practice. This was sometimes apparent in the interviews when we used library-centric vocabulary. For example, we used the term "materials" when referring to all the holdings in a library, whereas archivists use "archives" and "records" to mean the same thing. Sometimes our use of library terms temporarily confused some interviewees.

In round 2, we surveyed 8 of the original 13 institutions. We dropped five institutions from the study. Those institutions that remained through the two rounds are national archives. We eliminated one institution because they did not receive further funding for their program. A second institution needed to be eliminated because we could not re-establish contact with the respondent. In the third case, we determined that the institution did not still fit the parameters of our study. Rather than build on previous work, this participant was moving in different programmatic directions. The fourth institution turned out to be an outlier. In round 1 we had decided to look at one large research library program, but the program did not have the archival

[17] Grant McCracken, *The Long Interview* (Newbury Park, CA: Sage Publications, 1988): 12.

focus that characterized all the other institutions. For example, rather than focusing on strategies for born- and born-again digital materials, this institution was still focusing primarily on microfilming and digitization projects. Finally, in the case of the fifth institution, we deemed it appropriate to interview that person as a key expert in this round because the person has a broad and deep knowledge of digital preservation activities throughout the world.

Findings: The Survey

Eight institutions participated in the survey: five archives and three projects or programs. In round 1 we defined a project as "a specific undertaking or research endeavor, usually with special funding. Projects may take place within single, institutional programs, or at more than one site."[18] A program is "an ongoing set of services, around a common goal or activity, usually located within a single institution."[19] No libraries participated in this round. The institutions are based in Great Britain (3), the Netherlands (1), Australia (1), the US (2), and Canada (1).

Three years have passed from the first interview in round 1 to the last interview in round 2. The discussion follows the order of the questions in the survey questionnaire (Appendix 1). The respondents focused on their institutional collections rather than on their internal administrative or financial records.

Program and Policy

Since the first interview in 1999, seven respondents have reached the point of testing or evaluating preservation methods or techniques. All but one plan to disseminate results via Web sites, conferences, and publications. The remaining institution considers the information to be proprietary.

The preservation techniques or methods that the institutions are using or exploring include emulation (3), e-Permanence/Normalization (1), Persistent Object Preservation (1), and Migration (3). One project is exploring both migration and emulation.

The types of digital materials that are being preserved include business records, text documents, scientific data collections, databases, word-processed documents, Web sites, e-mail, image files, audio and video files, data sets from relational databases and spreadsheets, graphics (both vector and bitmap), and digital photography.

The responses to these questions focused mainly on programmatic issues rather than policy. We returned to questions about policy in the last section of the survey. We next wanted to learn whether institutions had progressed in determining criteria to choose a particular preservation technique, or if they provided updated information, whether that the decision was correct for their needs.

[18] Cloonan/Sanett, 96.
[19] Cloonan/Sanett, 96.

Specifics of Preservation Technique

Preservation techniques were chosen by institutions in a variety of ways, e.g., internal research and development, internal proprietary software system development, "software systems … developed based upon data handling systems, local collection management, digital libraries, web interfaces, and archival storage systems."[20] Other respondents indicated that the choice had not yet been made: "We are testing, as opposed to using, migration and emulation in the project. We are attempting to gain a deeper understanding both of cost issues and also of appropriateness of each technique for different categories of digital objects."[21] "In the test bed project, migration, emulation (not yet), and standardization (XML). No strategy has been chosen yet they are all under exploration in the tested project."[22] "We're pursuing the development of persistent archives as the primary technique."[23] One respondent is proceeding by "reference to best practice documents (e.g., "the Cornell University report on file formats")."[24] Evaluative data on the efficacy of the preservation method/model is pending for most of institutions surveyed.

Six respondents indicated that a factor in choosing the appropriate preservation technique for their institution included its potential effect upon the intellectual integrity (e.g., authenticity and reliability) of the digital material. Steps taken or planned to prove that the intellectual integrity of the digital material had not been compromised through the preservation process include "definition and documentation of the record's 'essence'; use of hashing techniques on XML source; [and] repeatability of the 'normalization' processes through the preservation of original bit streams."[25] One institution is "attempting to do this."[26] "Proof integrity (authenticity) [that] has not been compromised should be given by metadata about the preservation process and the management process in general. This should also be the case if in some way integrity has been compromised during the preservation process."[27] "In terms of the test sites, we prepared information packages with a view to protecting them from undocumented changes by restricting access to both read and write operations."[28] "Original bit stream preserved; all migrations will be checksum and regression validated. All migrations will be documented and new versions viewed as secondary manifestations."[29]

Responses indicate that steps have been taken to protect the intellectual integrity of the digital material and offer the reasons for choosing appropriate preservation techniques. These issues had surfaced during round 1, where respondents recognized the need to protect the intellectual integrity of the digital materials; but many of the projects/programs were then nascent and none of the respondents had yet taken those steps. Three years later, institutions have taken a variety of substantive steps to protect the intellectual integrity of the digital materials. This is indicative of the rapid progress made in this arena.

[20] Institutional Interview #5, July 30, 2001.
[21] Institutional Interview #4, November 14, 2001.
[22] Institutional Interview #3, September 1, 2001, in Washington, D.C.
[23] Institutional Interview #7, October 18, 2001.
[24] Institutional Interview #6, February 25, 2003. See Gregory W. Lawrence et al. *Risk Management of Digital Information: A File Format Investigation* (Washington, D.C.: The Council on Library and Information Resources, June 2000).
[25] Institutional Interview #1, September 11, 2001.
[26] Institutional Interview #2, December 12, 2001.
[27] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[28] Institutional Interview #4, November 14, 2001.
[29] Institutional Interview #6, February 25, 2003.

Cooperation

Institutional resources are usually stretched thin, so many institutions develop strategic internal and external alliances or collaborative relationships with public and private sectors. Primary areas of collaboration include research and development and testing.

Four institutions cooperated with other organizations to develop digital preservation initiatives, collaborating with archives, libraries, public companies, and other agencies. Collaborative efforts took place on local, national, and international levels. Work was not always distributed equally. The collaborative efforts evolved partially out of multiple funding resources; one respondent reported: "the project was funded by eight different agencies. Each project involved collaborations with other institutions."[30] Two respondents discussed various strengths and weaknesses attributable to the collaborative structure of the project. Strengths include complementary knowledge and/or experiences, multidisciplinary perspectives, breadth of efforts in data collection in digital libraries, data grids, virtual data grids, and development of the concept of persistent digital archives. There is tremendous energy in use of technologies from these research areas. Weaknesses cited include the multiple sets of competing standards for annotating information and knowledge, and dependency on the knowledge of others. However, the respondent noted that that might be an ongoing issue given the rapid developments in this area.[31]

Cooperative efforts first reported in round 1 continued in round 2. It appears that the strengths of the model outweighed weaknesses, or the collaborators are finding ways to resolve them. Regardless, with multiple research sites, similar research questions, and the need to be cost effective, the development of strategic alliances leading to collaborative relationships between institutions is in itself an effective strategy.

Staffing

We explored staffing issues because we wanted to learn about the roles that formal training and on-the-job training play in digital preservation**.** We wondered whether formal curricula are sufficiently able to identify and prepare students to acquire skill sets required for this type of work; we also wanted to learn whether employers relied on graduates of library, archival, and information science programs to staff their projects. Seven institutions responded that the staff involved in their preservation programs totaled 36.3 full-time and 42.1 part-time employees.

Five of these institutions did not consider the staffing to be sufficient, although one institution noted that staffing was sufficient for the current phase of their project. Regarding whether the respondents had the requisite skills required to work on the preservation of electronic records, several institutions answered both yes and no with comments. Required skills that were identified include:

- Document analysis, XML, and imperative programming skills; micro-appraisal skills; document modeling and definition;[32]

[30] Institutional Interview #5, July 30, 2001.
[31] Institutional Interview #3, September 1, 2001, in Washington, D.C., and Institutional Interview #5, July 30, 2001.
[32] Institutional Interview #1, September 11, 2001.

- Recordkeeping expertise, record information analysis, IT knowledge (and for the project, also project management, communication skills, project assistant, etc.);[33]
- Database, information management (XML, MIX), knowledge management (UML, DAML-S, model-based mediation);[34]
- "For the preservation of electronic records at a minimum you need [professionals in]three [areas]. One is an archivist who knows what the requirements of the work are and the nature of the objects they're dealing with. One is a computer specialist who can mount and maintain applications. And the other is an . . . 'archival engineer' who have a combination of knowledge from the archives side and from the IT side."[35]
- (1) Functional analysis: knowledge of an ability to perform functional analysis of electronic records systems; (2) Data management: knowledge of and ability to work with, data management concepts and systems; (3) Evolution of automation knowledge of and ability to recognize the various stages of the evolution of automation (hardware, computing environments, etc.); (4) Software: knowledge of and ability to work with various types of software; (5) Records management and document management systems: knowledge of and ability to work with records management and document management systems; (6) National Archives Practices and Approaches to ER: knowledge and ability to apply the institutions' practices and approaches; (7) Viewer and File editing software: knowledge and ability to use; (8) IM/IT Concepts: knowledge of ability to apply when working with e-records; (9) Trends and best practices in ER archiving: knowledge of trends and best practices in other archives and in the IM/IT industry.[36]

Four institutions noted that their organizations could provide staff that had the required skills, although one commented, "not in sufficient quantity for the future," and another commented that they "would need major staff development to locate, and train staff." The four respondents indicated that their preservation projects would draw staff away from other departments or projects.

We asked the respondents to describe the academic/professional background/job title/duties of the *full-time* preservation staff. Their responses follow:

- Professional archivists. Minimum is BA (Honours) in History and graduate studies in IT;[37]
- The internal persons involved are archivists and people with IT background; the contracted people: IT, computer science, archival theory;[38]
- Research interest in development of data, information, and knowledge management systems. Expertise in databases, information repositories, knowledge repositories, XML, UML, RDF, Topic Maps, ontology development;[39]
- IT/Computing/Engineering/Archives postgraduate level;[40]

[33] Institutional Interview #3, September 1, 2001, in Washington, D.C.
[34] Institutional Interview #5, July 30, 2001.
[35] Institutional Interview #7, October 18, 2001.
[36] Institutional Interview #8, March 1, 2002.
[37] Institutional Interview #1, September 11, 2001.
[38] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[39] Institutional Interview #5, July 30, 2001.
[40] Institutional Interview #6, February 25, 2003.

- Professional archivists with IT training; Master's degrees and one Ph.D. Will be looking for an academic computer scientist probably one holding a Ph.D.;[41]
- Training in history—others are geography, statistics, engineering, and computer science.[42]

Responses to the same question regarding *part-time* staff include:

- Information Science graduate diploma and an interest in preservation and long-term accessibility of Electronic Records;[43]
- Expertise in C, C++, logic programming, database applications;[44]
- Administrative/secretarial.[45]

The majority of respondents said that staffs receive training that supplements their project-based, on the job training. Types of training for staff include various continuing education courses offered by commercial providers, on-the-job training, participation in standards groups, review committees, conferences, and vendor collaborations. The frequency of staff updating their skills ranged from continually to more structured, quarterly training courses.

Timing of setting the preservation project in place ranged from less than three months (two respondents) to more than twelve months (four respondents). Three institutions expect to have the same staff complete the project that began, while two institutions do not. The respondents plan to contract out some of the major and/or minor components of the project. Contracted items include software development, hardware supply and installation, development of training programs, programming, and various minor non-core aspects of the project work.

In comparison with responses from round 1, where general staff skill sets included history, computer skills, and preservation and project management backgrounds, responses from round 2 were much more detailed with respect to skills expectations. This is consistent with evolving project needs of those projects that have progressed since round 1. Two points made in round 1 and round 2 are participants' concerns whether sufficient staff with required skills will be available in the future, and a comment from a participant in round 1 who has a background in history and computer science, that the professions are merging and that it is necessary to develop competencies for the merged professions. This comment may anticipate future needs, but the overall impression from the responses indicates that library schools, archival programs, and computer science programs should be developing strategic alliances to meet those needs.

[41] Institutional Interview #7, October 18, 2001.
[42] Institutional Interview #8, March 1, 2002.
[43] Institutional Interview #1, September 11, 2001.
[44] Institutional Interview #5, July 30, 2001.
[45] Institutional Interview #6, February 25, 2003.

Technical Questions

We asked participants to consider the following in order to explore the current status of preservation practices.

Is preservation carried out by the institution (in-house) and/or a commercial vendor/contractor?

Describe any pre-preservation preparation of records.

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

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

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

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

Most respondents carry out preservation activities in-house. Three respondents perform pre-preservation preparation of records. These activities include

- establishing a procedure for transfer from a government organization to the archive (in process);
- creation of an accessioning template, establishing digital object format and markup language, identification of the processing steps needed to convert to platform-independent representation; and,
- digital archive interface for loading records to be used by the archive and other government departments.

Of further interest is information that indicates the strong and weak points of the institutions' preservation methods or techniques. Respondents perceived their programmatic strengths as:

- Maintaining the integrity of preservation master once normalized; use of public domain/open source technologies; and almost certain lack of reliance on outside businesses over time.[46]
- Defining a process that supports the migration of collections onto new technology, in which the standards used to characterize the data also change.[47]
- Developing archival requirements rather than addressing technical problems.[48]

Respondents perceive their programmatic weaknesses as:

- Still at early stages of development.[49]

[46] Institutional Interview #1, September 11, 2001.
[47] Institutional Interview #5, July 30, 2001.
[48] Institutional Interview #7, October 18, 2001.
[49] Institutional Interview #1, September 11, 2001.

- Necessarily dependent on commercial standards becoming widely used for knowledge management, because the technologies will continue to change rapidly.[50]
- Not revolutionary in approach. Runs the risk of overheads [cost overruns] for migrations.[51]
- Dependent on the marketplace. "Most of what we are looking at doing is not currently available in the marketplace. The predictions are it will be there within a couple of years, that's one weak point. The other . . . is that to develop the system, we are going to need a lot more money."

Responses to questions regarding quality control indicate that such procedures range from 'still being worked out,' to internal audits in place.

Several procedures are reflected in the responses to the question of "how are you storing the electronic records that have been preserved?"

- On online media with optical disc and magnetic tape back-up.[52]
- Not yet happening.[53]
- We build a logical collection to store attribute values, aggregate digital objects into containers, and manage the storage of containers in archives through a data handling system.[54]
- Large-scale data servers.[55]
- It hasn't changed yet, but it will change. What we've looked for is reliable storage media. Most records being preserved aren't used most of the time so you want to write them to something that will last a while [be]cause it costs you a lot of money to copy it to new media. It's beginning to look like copying to new media can actually save you money, thanks to Moore's Law.[56] So what we're looking at is a new equation which says, "it's not so much that I want it to last for 10 years versus 5 years, I want it to be reliable for however long I'm going to keep it and that length of time factors in[to] how much it cost me to keep it on that medium versus to migrate it to a newer medium which is denser and faster and less expensive.[57]

Between round 1 and round 2, new technological resources have become available for data storage. Programmatic strengths and weaknesses discussed in round 2 naturally illustrate a greater level of detail than provided in round 1. The needs for widely accepted standards for knowledge management and quality verification, as well as a lack of sufficient financial resources, were points made in both survey rounds. Implementing satisfactory quality control procedures remains a goal for many institutions. The lack of progress on this issue may be linked to the need for standards, and to further research and development into the technology necessary to support quality control processes. This lack may also be attributable to the constant evolution of hardware and software in the field.

[50] Institutional Interview #5, July 30, 2001.
[51] Institutional Interview #6, February 25, 2003.
[52] Institutional Interview #1, September 11, 2001.
[53] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[54] Institutional Interview #5, July 30, 2001.
[55] Institutional Interview #6, February 25, 2003.
[56] For a discussion of Moore's Law see Gordon E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* 38 (April 19, 1965), or http://www.intel.com/research/silicon/moorespaper.pdf.
[57] Institutional Interview #7, October 18, 2001.

Cost Activities

Increasingly, participants at conferences focus on the need of institutions responsible for preserving records, to determine potential costs to preserve electronic records over time.[58] Issues include methods to determine costs, selecting appropriate cost elements, the use of cost-benefit versus risk-benefit analysis, development and analysis of costing frameworks and cost models, and the integration of this information with institutional decision-making strategies. Discussions lead into the need for policy development in this area, largely because of the need for institutional support over time. Based on responses in round 1 of our survey, and the increasing interest in this topic, we broadened the scope of our questions about cost activities, and deepened them, hoping to learn what progress institutions had made in developing cost figures and policy related to the preservation of electronic records. We grouped questions into three broad categories: value, budget, and insurance. Policy is considered in the last section of this survey.

Value

When we asked whether respondents could estimate the monetary value of the records in the archive, only one respondent said yes. One respondent commented that in a government archives the records have evidential, business, and cultural values. We then asked the respondents who answered 'no' whether there is a problem measuring the value of the records in the archive. The discussion ranged from philosophical to accounting. There was no consensus on how to value records, or whether they should be valued in a monetary sense.

When asked what the estimated value of the archive is, one institution responded: "The cost of acquiring the data."[59] Asked how that value was calculated, two respondents commented:

- For data created by a person lost, the value is the effect it has on whatever you're trying to do with the data in the future. The value of the data is the purpose you're going to use it for in the future.[60]
- There are cases where one could absolutely assign a specific value because the existence of a given record which the government will certify the authenticity [of] can make a court case-make or break it. But that depends on the use to which you're putting it.[61]

In both of these responses, the value of the records in the archives was closely connected to function of the archives.

Budget

We asked whether respondents have an annual budget for digital preservation activities. Five respondents do, one does not. One respondent indicated that there is not one identified budget – parts of various budgets are used. Annual budget amounts (in U.S. dollars) ranged from $300,000 to $3 million, with additional funding requested by one respondent.

[58] Two recent examples include the National Archives/ICA Conference, Kew, England, April 2003 and the ARSAG Conference, Paris, France, May 2002.
[59] Institutional Interview #5, July 30, 2001.
[60] Institutional Interview #5, July 30, 2001.
[61] Institutional Interview #7, October 18, 2001.

When asked, "If you do not have a regular budget for digital preservation, or if the budget is demonstrably insufficient, does this create a risk that you can measure?" Two institutions replied in the negative. One commented, "No, but there is a huge risk, of course, for losing electronic records."[62]

Three institutions indicated that problems could be resolved with additional funds. Another respondent pointed out that "There are always problems that can be resolved with more money, but equally there are always problems that will exist irrespective of how many $$$ you throw at them." [63] We also asked what the impact would be if these problems persist. Responses were similar:

- The impact would be that the Archives cannot properly preserve electronic records.[64]
- It's hard to quantify the value in future. The cost of the labor will still be there. The major challenge is finding funds to apply the technology to new disciplines.[65]
- Loss of accessibility to the records.[66]

When asked whether the annual budget was sufficient for routine digital preservation activities, two institutions responded yes, two responded no. One commented that their institution did not know yet. Four institutions responded that their current budget could fund a project to preserve electronic records. One institution did not responded, while another was uncertain. Amounts allocated for this purpose ranged from $100,000 (US) to support the preservation of an electronic records repository to $250,000 (US) for research and development. These institutions considered the amounts to be sufficient for the activities. One respondent commented that he/she doesn't know yet.

We also asked institutions to estimate the annual costs to preserve their records; as expected, responses varied. We asked them to include staff, equipment, space, energy, and other related costs.

- $250,000 (US)[67]
- No idea.[68]
- $100,000 (US)[69]
- Unknown[70]
- We have not done the estimate. In addition, some fixed costs are shared, and therefore difficult to assign.[71]
- Private[72] [the information is classified]

[62] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[63] Institutional Interview #1, September 11, 2001.
[64] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[65] Institutional Interview #5, July 30, 2001.
[66] Institutional Interview #8, March 1, 2002.
[67] Institutional Interview #1, September 11, 2001.
[68] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[69] Institutional Interview #5, July 30, 2001.
[70] Institutional Interview #7, October 18, 2001.
[71] Institutional Interview #8, March 1, 2002.
[72] Institutional Interview #6, February 25, 2003.

One institution indicated how they calculated the estimate: "We have the luxury of additional projects that are developing the underlying technology (data handling system, information catalog management system, knowledge characterization). We can build a collection and archive it for the labor cost to support automated ingestion of the collection."[73]

Additionally, we asked what categories are included in the current preservation/digital preservation budget to get a sense of the cost categories institutions related to the cost of digital preservation activities. We probably should have confined the question to accounting categories because institutions interpreted the question in different ways:

- All archival electronic records taken into custody.[74]
- Regarding the Archives, only one FTE and a relatively small budget is available.[75]
- Research on data management, information management, and knowledge management, encapsulated as data handling systems, distributed collections, digital libraries, data grids, and persistent archives.[76]
- Research and Exploratory Development; progressive deployment, money spent for short-term solutions to meet pressing operational needs. The big category is systems development, and then there is another category we're calling business development transition management, which is "how do you bring the organization forward so it can operate in a digital environment?"[77]
- There is no budget, per se, however the funds we spend cover: technical analysis of records during appraisal, leading to the specification of Terms and Conditions of Transfer, initial processing after transfer (conversion to standard formats), some description, making of second copies of every record (different physical storage format), keeping preservation metadata, ongoing physical format conversions (e.g., nine-track tape to 8 mm. cassette and digital linear tape), and physical condition surveys.[78]
- Private[79] [the information is classified]

Five institutions responded to the question "When you estimate the costs to preserve electronic records, do you expect the categories to be the same or different from those in your current preservation/digital preservation budget?" Four said that they do not expect it to be different. One considered that information to be private.

Insurance

When asked whether the institutional holdings were insured, four institutions responded that they were not, one said they were. Two government archives commented that their governments carried their own insurance; thus the collections are self-insured. None of the respondents which have uninsured holdings plans to obtain insurance. Reasons for not doing so include:

[73] Institutional Interview #5, July 30, 2001.
[74] Institutional Interview #1, September 11, 2001.
[75] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[76] Institutional Interview #5, July 30, 2001.
[77] Institutional Interview #7, October 18, 2001.
[78] Institutional Interview #8, March 1, 2002.
[79] Institutional Interview #6, February 25, 2003.

- The records, by definition, are irreplaceable.[80]
- The _____ government mostly does not insure things. Furthermore insurance does not make much sense. Records are unique, so in case of loss or a disaster they will be lost anyway. The only insurance that could be taken is to make copies and put them in other places. This policy is followed incidentally and might become structural in case of digital records.[81]
- We are still a research effort, and have no authorization to spend research funding on insurance.[82]
- It has generally been considered that, except in a few special cases, commercial insurance would not provide value for money for the government because the size and range of its business is so large that it does not need to spread its risks, while the value of claims met is unlikely to exceed its premium payments.[83]
- As a matter of policy, the Government of _____ does not insure itself.[84]

When asked how the value of uninsured holdings was determined, none of the institutions responded.

Many cost issues appear in both rounds of the survey. As projects and programs mature, substantive data on actual costs to preserve electronic records should emerge. This topic should be revisited periodically.

Records Organization and Provenance

We asked for an update on previous responses to the round 1 survey. The questions we asked were:

How are preserved records organized?

The responses indicated there was no change between round 1 and 2; i.e., records are organized by the originating department, functionality, an intellectual control system, or archival principles, as part of the *fonds* to which they belong.

How is provenance respected?

Again, responses provided greater detail, but essentially there was no change between round 1 and round 2. The primary methods used include documentation and record organization processes.

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

Participants indicated that electronic records are subject to the same accessibility restrictions as other government records. These restrictions differed in scope among institutions. Controls range from systemic to metadata to access control lists on each object or a combination.

[80] Institutional Interview #1, September 11, 2001.
[81] Institutional Interview #3, conducted September 1, 2001, in Washington, D.C.
[82] Institutional Interview #5, July 30, 2001.
[83] Institutional Interview #6, February 25, 2003.
[84] Institutional Interview #8, March 1, 2002.

Under the rubric *Preservation Plans, Actions, Procedures, and Policies*, we asked several questions. This area evolved from round 1 of the questionnaire, where it was evident that policy development was lagging behind technological advances. We wanted to learn whether there was less of a gap three years later. The questions we asked were:

Does your institution have a digital preservation plan?

To this question we received eight respondents; three did and three did not have such a plan, while two did not answer the question. One participant's plan was undergoing revision at the time of the survey.

If not, does not having a digital preservation plan create a risk you can measure? Will this risk have an impact on the institution or archive?

A noteworthy response was, "A plan is in preparation at _____. As long as there is no plan the risk will be that preservation will be *ad hoc*, inconsistent and not embedded in the organization."

Of the institutions with digital preservation plans, the plans are readily available to archive/institution staff, organizational management, and archive stakeholders. Respondents reported that staff, including management, reviews plans regularly.

Can the institution's administration use the plan to understand how strategies to preserve electronic records fit into the operations of the archive or institution?

Five participants responded, four answering positively, one negatively.

Are actions and/or procedures being formulated into policy?

Two institutions responded that they were. There were six non-responses.

Do you have a general preservation policy that includes records in electronic form?

Of the six responses, five participants answered affirmatively.

If not, do you have a policy for reformatting, refreshing, migrating, and emulating data to newer technological platforms?

There were two responses to this question; one institution does have a policy, the second did not.

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

There were only two responses to this question. One participant was quite specific as to the scope of the policy: "The major policies are the specification of: standards to use for markup languages; migration policies (how frequently to track technology updates); access standards (what mechanisms to use for discovery of data); and, conformance to emerging persistent archive standards."[85]

[85] Institutional Interview #5, July 30, 2001.

A second respondent said that the policy was included in annual business plan targets.[86]

In round 2, five of eight institutions (63%) surveyed have a general preservation policy that includes records in electronic form. In round 1, only 3 of 13 had such a policy (23%). Because the number of institutions in round 2 is smaller than those surveyed in round 1, this information should be considered only as a potential indicator of an emerging trend.

Findings: The Key Informant Interviews

The key informant interviews provided many perspectives on current preservation programs, activities, and research. That they also represented broad, expansive, and even utopian perspectives is illustrated in the responses quoted, below.[87]

The first question was asked in round 1: "How do you define preservation and has your definition changed over the years?" In round 1 we reported a shift in the definitions away from the artifactual approach that predominated for paper-based records, to the view for digital records that today the aim is to preserve the ability to reproduce a record.[88] From the key informant interviews, a more broad-stroke perspective emerged. Two of the interviewees (Deanna Marcum and Don Waters) emphasized that preservation is stewardship. "I keep adding pieces to my original understanding of preservation and I don't think of it in terms of '500 years' or 'techniques' but rather more in terms of stewardship: what body of materials are we as libraries and archives responsible for from society's perspective? . . . So I do not have any changing definitions of preservation. But I think that the stewardship rule does not change" (Marcum). Don Waters similarly said, "I see preservation as a function of stewardship and custodianship."

An archival scholar and educator, Sue McKemmish, said, "Speaking of 'recordness' and recreating 'recordness' in a digital environment: Recordness relates to the qualities that make something a record, like having a fixed content and always being re-presented in the structure or form in which it was born, along with being able to recreate the context in which it was created. It also entails being able to demonstrate and track what has happened to the record over time; this includes what recordkeeping management processes have been applied to it—being able to have an audit trail of that. . . . You can talk of preserving content, preserving the structure, preserving the context, and continually linking the records themselves to the related records of the recordkeeping processes. There may be many implementation solutions in terms of the way that digital preservation is carried out."

Ian Lancashire suggested a more utopian definition: everyone must be a preserver in the digital world. "I would like to make a pitch for preservation at a local level, at the individual level. I'd like each individual to be able to preserve his or her digital materials and contribute to some kind of archive, a preservation institution."

[86] Institutional Interview #6, February 25, 2003.
[87] In this section we identify the key informants by name because we specifically asked for their personal views.
[88] Cloonan and Sanett, 84-87.

Under "Staffing, Education and Training," we wanted to know whether the interviewees felt that LIS programs are up to the task of preparing students to work with (and preserve) digital records. We received a wide variety of responses, with Anne Kenney and Abby Smith asserting that the most important skill to have is critical thinking. They both felt that a good liberal arts education provided that grounding. Kenney also emphasized the importance of archival training: "When we started to do the work on digital image quality, my archival experience was much more important to me than any technical knowledge. It was that profound love of the document that kept me engaged in terms of what the digital surrogate says in relationship to the source document itself, not what the technology was capable of doing. Being able to make those basic concepts relevant in a technical realm should be an important part of formal education."

Howard Besser also weighed theory versus practice when he said, "The most . . . critical part of educational programs is that people learn concepts first and practices as instantiations of concepts."

Ross Harvey said, with respect to education for digital preservation, "it is not the techniques that are particularly different, it is the way of thinking about how they are applied—which is where the change happens."

Almost everyone felt that LIS programs were important venues for basic professional education, but that increasingly we also need people with backgrounds in computer science, business, legal, and other related fields to deal with all aspects of digital records. Paul Conway observed, "The cultural materials community has a 'preservation mandate' at the heart of its professional ethics. It has the richest, most complex, and varied conglomeration of materials that need to be preserved in either traditional or digital formats."

Some interviewees, particularly Philip Bantin, emphasized that we also need crossover skills from business, computer science, law, and engineering. This can happen in several ways: LIS students at some schools can earn joint master's degrees, we can recruit students into librarianship and archives who already have these backgrounds, and we can provide many continuing education opportunities for professionals.the

In the section on "Preservation Planning, Actions, Procedures, and Policies" we first asked the interviewees to identify three key preservation problems that affect born digital materials." Although many of the respondents initially said something like, "the usual three," a wide variety of answers emerged, as is shown in Appendix 5. This diversity points to the many challenges presented by digital assets.

Phil Bantin summarized the problems cogently when he said that there is a "lack of emphasis on management issues that will make preservation routine, systematic, and implemented at the right times for the right kinds of things."

The second question, "What might be some advantages to developing national preservation policies?," turned out to be something of a national litmus test. Almost to a person, Americans were skeptical of the efficacy of national policies. Nancy Elkington, who spent several years working for the Research Libraries Group in the UK, astutely noted that, "While national policy

seemed to work as a driver in [the UK], Americans are not a nation prone to cultural policy making. [Instead] we have focused on guidelines."

Marcum and Friedlander indirectly allude to this ambivalence when they write, "Part of the problem reflects our uncertainty in this country about who is responsible for preservation of library resources. In other developed countries, a national library has responsibility for acquiring and preserving a country's published output. Our Library of Congress serves as a national library in some respects but has no universal preservation charge for the nation—[even though] it has recently been assigned the responsibility . . . to support long-term preservation of digital content through the National Digital Information Infrastructure and Preservation Program (NDIIPP)."[89]

Perhaps as a consequence, preservation activities in the United States have tended to be de-centralized. Instead, Americans have established programs, guidelines, and standards through professional organizations, consortia, and other cooperative activities.

A couple of the American interviewees did cite some possible advantages, such as that "it would raise the visibility and awareness of digital preservation issues" (Don Sawyer) and that it could "form a standard of minimum practice" for institutions (Evelyn Frangakis).

The international respondents identified two additional advantages of having national preservation policies: there would be no preservation have-nots (Lancashire) and national policies could be aimed at coordination and optimization (McKemmish).

But they also identified some potential disadvantages. McKemmish noted that there are downsides to national policymaking and that one could "lose the leading edge . . . like the committee that designs the horse and winds up with the camel." Harvey speculated that such policies might cause some programs to fold.

We also asked "Who do you see as the prospective future users of digital materials?" Every respondent indicated that everyone is a prospective user.

We asked the interviewees about cost in its broadest interpretation, as well as its role in decision making. Regarding this broad view, most respondents felt that inaction may bring about inestimable loss of our cultural heritage.

Regarding cost in decision-making, Anne Kenney noted that:

> We measure dollars and we measure production, but we don't measure how well the program meets the needs of an institution. I get concerned when people look at cost and start making assumptions about how successful a program is. With respect to "cost benefit," you must look at it in terms of the cost of *not* doing something, the cost of doing it, and the cost saved by doing it as part of a consortial effort of contributing institutions. We must be wary of how we define "cost assessment" in terms of decision making. Cost is reality training, first and foremost."

[89] Deanna Marcum and Amy Friedlander, "Keepers of the Crumbling Culture: What Digital Preservation Can Learn from Library History," *D-Lib Magazine* (May 2003). http://www.dlib.org/dlib/may03/friedlander/05friedlander.htm/

And in a similar vein, McKemmish asked,

> How do you justify putting resources into preservation of digital records, how do you demonstrate the benefit to the community, and how do you provide some kind of notion of what it would cost the community not to have that happen?

Other perspectives emerged:

> We are at the birthing stage of developing costs. If we do things right, we won't need an assignable cost—it will just be a part of what we do, like mopping the floor. (Elkington)

> Cost can be an impediment to moving forward because people use cost as an excuse for not acting. (Waters)

> A lot of the costs in the long-term maintenance issues for digital materials gets back to [the costs of] continuous appraisal. (Cox)

And similarly,

> Some issues are the same as in the paper world: what do you select, how do you keep it? Also, there is the cost of maintaining that information over a period of time. (Harvey)

Informants said we must not merely find short-term solutions. This perspective is consistent with survey results.

One of the two wrap-up questions was, "What are your great concerns regarding the preservation of records?"

The responses to this question mentioned lack of public awareness, inadequate funding, rights management issues, and social and political issues. This overlapped with our earlier question about key challenges, but here, the interviewees tended to place their concerns into a broader social context. For example:

> One of my concerns is the will to preserve in a disposable society. Benign neglect is not an option in a digital world. (Frangakis)

> We must engage the public in such a way that they understand that preservation is a social and cultural activity. (Smith)

> Migration makes archival appraisal a continuous process; every time you have to do a migration, you have to reappraise. (Cox)

> One of the things we are concerned about is the records in our custody: can we certify that they are what we say they are? The chain of custody is important. One of the biggest concerns is insuring the integrity of the record. (Jones)

> Vast amounts of money are going to preserve government records, yet no two agencies have the same practices. (Elkington)

Finally, Waters described what he identified as 'public good' versus 'free rider' issues. ". . . in the digital medium, where it's very easy to copy things from one place to another, it . . . brings to the forefront the whole 'public good' problem that preservation represents anyway, and that is, perhaps, something the culture at large benefits from, if a single institution takes responsibility for preserving an object. . . . And yet it's not always the case that the preserving institution can get the financial support of the public that benefits from it. So it is often a huge 'free rider' problem and it leads to institutions saying, 'Why should I bother? Let someone else take care of it."

For the final question we asked the interviewees to discuss current projects. Since the projects were institution specific, the answers did not yield generalizable results for the present study.

Conclusions

The data collected from the surveys show that since round 1:

- There is still a lack of consensus on preservation strategies and most institutions are continuing to explore one particular type of preservation strategy.
- Institutions are beginning to address cost issues by seeking models or creating new ones.
- An increasing number of institutions are developing policies or seeing the need to develop them.
- Respondents have a clearer idea of their staffing needs over the next three to five years than was indicated by the round 1 participants.
- Nearly half of the institutions surveyed in round 2 have secured funding for R & D—through budget increases, government appropriations, collaborations, and grants.
- Institutions are actively continuing to develop strategic alliances and collaborations that are largely centered on project-based R & D.

The key informant interviewees identified many digital preservation challenges ranging from technical to social, political, legal, and educational issues (see Appendix 8). The key informants stressed the need for technical solutions such as more standards, and a greater focus on issues such as interoperability, scalability, and automated metadata. At the same time the informants lamented "the muddy waters of rights management" (Elkington) as well as the lack of awareness of digital preservation issues among the general public. Finally, they identified the need to educate people to address these issues. The interviewees said that ongoing continuing education is crucial, but there were different ideas about how to educate potential professionals. Some felt that with a solid liberal arts education, one could learn "on the job." Others felt that there was a

need for more professional librarians and archivists as well as computer scientists, economists, lawyers, and managers. Those who felt that a master's degree in library, archival, and information science was crucial also believed that current programs are not addressing preservation needs adequately.

Taken together, the surveys and key informant interviews indicate that

- Both the nuts-and-bolts institution-specific approach of the survey respondents and the global perspectives voiced by the key informants are crucial to an understanding of the full range of digital preservation issues.

- Both groups are concerned with issues relating to the authenticity and reliability of records in electronic systems. Survey respondents are trying to incorporate these concepts into their institutional procedures while the key informants are trying to create standards, models, and national and international initiatives that respect authenticity and reliability.

- Both groups expressed the need for more research in a variety of areas.[90]

Figure 2 highlights secondary themes that came out in the survey and key informant interviews.

Figure 2. Secondary Themes
(Not specifically asked in the questions, but revealed in the interviews)

Survey Questionnaire	Key Informant Questionnaire
Collaborations/Strategic Alliances	**Collaborations (national, international)**
Authenticity and Reliability (a.k.a. Trustworthiness)	**Authenticity and Reliability (a.k.a. Trustworthiness)**
Need to develop standards	**Ongoing need for standards** **Ongoing need for interoperability**
Need for research in technology, processes/procedures	**Need for more research on** • **Best practices** • **Use of archival collections** • **Preservation strategies (systems and technologies)**

There was one unexpected outcome to the research. We did not ask the key informants about professional or theoretical influences in their work, yet everyone cited influential readings or scholars. A future strategy might be to study the impact of specific influences on the approaches that key informants bring to their work. As far as this study went, this was secondary data that emerged through data analysis, and we chose to include it as appendix 7.

[90] A recent paper does an excellent job of identifying areas that researchers need to address. See Margaret Hedstrom and Seamus Ross, *Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation.* Prepared for the National Science Foundation's (NSF) Digital Library Initiative and the European Union under the Fifth Framework Programme by the Network of Excellence for Digital Libraries (DELOS), 2003.

What do we need to know? What are the implications of our findings?

During the three-year period of our two surveys, projects progressed or were completed and the practitioners interested in the challenges of long-term preservation of electronic records now have results to review. Participants in this round represented most of the major archival funded projects. Institutions that are actively involved with seeking solutions to these problems were represented in this survey. The wealth of information participants provided illustrated the rapid technological and intellectual progress made during the last three years. There is now more thoughtful literature available on problems related to long-term digital preservation than there was three years ago. Also, various camps have solidified behind particular preservation strategies. In the short time between surveys, we moved from managing (and needing to preserve) mostly flat files, to dealing with files with links, to multi-media, dynamic, experiential, and interactive records.

As we stated at the beginning of this study, the purpose is to identify approaches to and ideas about digital preservation. We reported the findings of round 2 of our study, in which we continued to document extant practices as well as expert opinions on the digital preservation dilemma. We explored some of the known roadblocks to long-term digital preservation, such as methods for preservation, costs, intellectual property rights issues, and selection and appraisal for preservation.

We particularly note that there is not as yet a consensus on a single preservation strategy. Should we expect it or want it? We suggest that an area of further research might be an exploration of several forms of preservation strategies within one project--a suite-of-tools approach. It may be that one type of preservation strategy is more effective for particular types of records than another. Or a particular type of preservation strategy may be more cost-effective to use than another for a particular institution's needs--and for the type of records they hold, they can make choices as to strategies. At some point we as a community must find answers to these challenges and take the opportunity to compare the performance of various preservation strategies and their costs across types of records, as an institution would have to do in order to make choices appropriate to its operations and mission. In an ideal world, the pragmatism of the survey respondents and the broad perspectives of the key informants should function to support each other's efforts. This may become the next step toward taking research closer to real-world applications, which, after all, is the heart of issues we have explored.

At the end of round 2, we find that preservation activities for digital assets can be *ad hoc* or inconsistent. It is promising that some of our survey institutions are thinking strategically. In round 3 we will take a close look at the strategies employed at two of the institutions surveyed.

March 29, 2004

Appendix 1: InterPARES Questionnaire on Preservation Strategies for Electronic Records, Round 2

NOTE: Please answer only those questions that apply to your institution/project/program. Some questions ask for an update on information since Round 1. If those apply to you, please supply new/additional information only.

1. *Date of Interview:* _____

1.1. Name of Interviewee:

_____

1.2. Title of Interviewee:

_____

1.3. Name(s) of Interviewer(s):

_____

A. About the Archive/Institution(s) [please update if there have been any changes]

2. Name of Institution or Project:

_____

2.1. Person (s) responsible for the program (Name (s) and Title (s)):

_____

2.2. Address:

_____

_____

2.3. Telephone Number: _____

2.4. Fax Number: _____

2.5. E-mail: _____

B. Program and Policy

3. Describe in broad terms what methods or techniques you are exploring or using for digital preservation since our last interview.

_____
_____
_____

3.1. Describe the digital materials your institution is preserving since our last interview.

_____
_____
_____

3.2. Has the program or activity reached the point of either testing or evaluating any of the methods or techniques you are using since our last interview? If so, what are the results to date?

_____
_____
_____

3.3. Will you be disseminating the results? How?

_____
_____
_____

3.4. How do you define "original"?

_____
_____
_____

3.5. How do you define "copy"?

_____
_____
_____

C. Specifics of Preservation Technique [Please update]

4. Which preservation techniques does your program use? _____

4.1. How was each technique selected?

_____
_____
_____

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

_____
_____
_____

4.3. In selecting the preservation technique, have you considered what its effect might be upon the intellectual integrity (e.g., authenticity and reliability) of the digital material? θ YES
θ NO

4.4. 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.

_____
_____
_____

4.5. Is there evaluative data on the efficacy of this preservation method/model? Please describe.

_____
_____
_____

D. Selection for Preservation (Deferred to Round 3)

Cooperation [please answer only if you have new partners/collaborators since Round 1]

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

6.1. If so, which? (Check all relevant)

θ Archives θ Libraries θ Public companies
θ Museums θ Other (please explain)

6.2. Is your cooperation θ National θ International θ Local
θ Shared facilities θ By institutional type
[churches, labor unions, etc.]

6.3. How is the work distributed?

θ Equally θ Work distributed in a different way

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. How many staff are involved with the program?

θ Full-time? How many? _____
θ Part-time (FTE)? How many? _____

7.1. In your estimation, is the staff

θ Sufficient
θ More than sufficient
θ Insufficient

7.2. Have you identified the skills required to work on the preservation of electronic records?
θ YES θ NO

7.2.1. If yes, please describe the required skills.

_____
_____
_____

7.3. Can your organization provide staff that have the required skills?
θ YES θ NO

7.3.1. Will your project draw staff away from other departments or projects?
θ YES θ NO

7.4. Describe the academic/professional background/job title/duties of the preservation staff.

Full-time staff:

_____
_____
_____

Part- time staff:

_____
_____
_____

7.5. Does your staff receive training only in the project? θ YES θ NO

7.6. Does your staff receive outside training? θ YES θ NO

7.7. Please describe the training the staff receives.

_____
_____
_____

7.8. Please describe how and how often the staff updates skills.

_____
_____
_____

7.9. Can you estimate how long the project will take to set in place? If so, please estimate the approximate time.

θ Less than 3 months
θ 3-12 months
θ More than 12 months

7.10. Can you expect to have the same staff that begins the project complete the project?
θ YES θ NO

7.11. Does your institution plan or need to contract or obtain outside assistance for

θ Minor components of the project
θ Major components of the project
θ The complete project

Please explain any checked items.

_____
_____
_____

Technical Questions

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

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. Cost Activities

9. In some respects, money spent on digital preservation efforts, is an investment an organization makes to ensure continuing access to the information. In this sense, the value of the records, or the cost of not having the records should increase with time. At this time, can you estimate the monetary value of the records in the archive?

θ YES θ NO

9.1. If no, is there a problem measuring the value of the records in the archive?

_____
_____
_____

9.2. If yes, what is the estimated value of the archive? $_____

9.2.1. How did you calculate this value?

_____
_____
_____

9.3. Do you have an annual budget for digital preservation activities? (Digital preservation can include but not be limited to migration, scanning, emulation, refreshing, metadata creation and related activities.) θ YES θ NO

9.3.1. If yes, what is your annual budget? $_____

9.4. If you do not have a regular budget for digital preservation, or if the budget is demonstrably insufficient, does this create a risk that you can measure?
θ YES θ NO

9.4.1. Are there problems that could be resolved with additional funds?
θ YES θ NO

9.4.2. If these problems persist, what impact could they have on the archive?

_____
_____
_____

9.5. Is your budget sufficient for routine digital preservation activities?
θ YES θ NO

9.6. Can your current budget fund a project to preserve electronic records?
θ YES θ NO θ Uncertain

9.6.1. If yes, enter the amount you can allocate for this purpose. $_____

9.7. In your estimation, will these funds be
θ Sufficient? θ Insufficient?

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

_____
_____
_____

9.8.1. How did you calculate your estimate?

_____
_____
_____

9.9. What categories are included in your current preservation/digital preservation budget?

_____
_____
_____

9.9.1. When you estimate the costs to preserve electronic records, do you expect the categories to be the same or different from those in your current preservation/digital preservation budget? Please explain.

_____
_____
_____

9.10. Are the holdings of the archive/institution insured? θ YES θ NO

9.10.1. If no, are you considering obtaining insurance? θ YES θ NO

9.10.2. If no, please describe objections to obtaining insurance?

_____
_____
_____

9.11. If the holdings of the archive/institution are insured, how was the value of the holdings determined?

_____
_____
_____

9.11.1. What is the cost of the insurance coverage? $_____

9.11.2. Does the cost increase annually? θ YES θ NO

9.11.3. By how much? $_____

I. Records Organization and Provenance [Please update]

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 [Please update]

11. Describe record-keeping for the preserved material.

_____
_____
_____

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

11.1.1. If yes, which one? _____

11.2. How is metadata used to describe preserved materials?

_____
_____
_____

K. Access to Preserved Records [Please update]

12. Are the preserved records available θ only on-site θ only within the institution θ through a website θ other (specify)

_____
_____
_____

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

_____

12.2. Data users may or may not be stakeholders in your archive or institution. Stakeholders are interested individuals or groups who have a voice in the various aspects of the institution's implementation of the preservation project. Are data users stakeholders in your institution?
θ YES θ NO

12.2.1. If yes, can you describe how your data users are involved in preservation decisions?

θ Constituents are heavily involved.
θ Constituents are routinely consulted.
θ Constituents contacted only as needed.

L. Charges for Access (Defer to Round 3)

M. Reproduction and Copyright (Defer to Round 3)

N. Preservation Plans, Actions, Procedures and Policies [Please update]

15. Does your institution have a digital preservation plan? θ YES θ NO

15.1. If not, does not having a digital preservation plan create a risk you can measure? Will this risk have an impact on the institution or archive?

_____
_____
_____

15.1.1. If yes, has the plan been thoroughly reviewed by management? θ YES θ NO

15.1.2. If yes, is the plan

θ Readily available to archive/institution staff?
θ Readily available to the organizational management?
θ Readily available to archive stakeholders?
θ Regularly reviewed?

15.2. Can the institution's administration use the plan to understand how strategies to preserve electronic records fits into the operations of the archive or institution?
θ YES θ NO

15. 3. Are actions and/or procedures being formulated into policy? θ YES θ NO

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

15.4.1. If not, do you have a policy for reformatting, refreshing, migrating, and emulating data to newer technological platforms? θ YES θ NO

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

_____
_____
_____

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

_____

_____

_____

06/28/01

Appendix 2: Key Informants/Experts, InterPARES Survey, Round 2

Questions

Program and Policy

1. How do you define the term *preservation?* Has your definition changed over time?

Staffing, Education and Training

2. In North America, at present, only a few Library & Information Science/Archival programs are offering regular courses in preservation. Where and how will we train digital preservation specialists? What type of qualifications do we expect digital asset managers to have? Will we need to draw from other fields?

Preservation Planning, Actions, Procedures and Policies

3. Identify three key preservation problems that affect born digital materials. What are some barriers to overcoming them?

4. What might be some advantages to developing national preservation policies? Disadvantages?

What components should be included in promulgating national preservation policies?

Access to Preserved Records

5. Who do you see as the prospective future users of digital materials?

Cost Activities

6. What is the Cost of preservation in the digital age?

7. What do you see as the role of cost in decision making?

8. Sources of funding; constraints on funding; how funding affects preservation policy or practices?

Concluding Questions

9. What are your greatest concerns regarding the preservation of records?

10. Is there anything you want to add to what you've already said? Is there anything that you would like us to know about your work vis-à-vis preservation?

Appendix 3: InterPARES Preservation Survey Legend, Round 2

Contact Interviewed	Completed
1. Simon Davis, National Archives of Australia (**A) (EP/N)**	September 11, 2001
2. Stewart Granger, CAMiLEON **(P) (E)**	December 12, 2001
3. _____, Ministry of the Interior, Netherlands **(A)(E)**	September 1, 2001
4. _____, Univ. of Leeds, CURL, CEDAR **(P) (E)**	November 14, 2001
5. Reagan Moore, San Diego Supercomputer Center **(P) (POP)**	July 30, 2001
6. David Ryan, National Archives, England **(A) (M)**	February 25, 2003
7. Ken Thibodeau, NARA **(A) (M)**	October 18, 2001
8. _____, National Archives of Canada, now part of Library and Archives Canada **(A) (M)**	March 1, 2002

Type of Organization:
A = Archive (5)
L = Library (0)
P = Project/Program (3)

Primary type of Preservation strategy in use or being researched:
E = Emulation (3)
EP/N = e-Permanence/normalization (1)
POP = Persistent Object Preservation (1)
M= Migration (3)

Revised May 25, 2004

Appendix 4: Key Informants Interviewed, Round 2

(arranged chronologically)

	Interviewees	Date	Location
1.	Abby Smith/Deanna Marcum/ Anne Kenney	8/29/01	CLIR, Washington DC
2.	Norvell M. M. Jones	11/05/01	NARA, Washington, DC
3.	Evelyn Frangakis, Don Sawyer	11/05/01	NAL, Washington, DC
4.	Philip Bantin	11/27/01	via telephone from UCLA
5.	Paul Conway	8/23/02	SAA, Birmingham, AL
6.	Nancy Elkington, Robin Dale	1/28/03	RLG in NYC, with Robin via telephone
7.	Don Waters	1/29/03	the Mellon Foundation, NYC
8.	Richard Cox	1/31/03	via telephone from AMIGOS, Dallas, TX
9.	Michael Lesk	1/31/03	first via email 11/5/01, then follow up phone interview from AMIGOS, Dallas, TX
10.	Ian Lancashire	2/10/03	InterPARES meeting, Vancouver, B.C.
11.	Howard Besser, James Turner	2/11/03	InterPARES meeting, Vancouver, B.C.
12.	Sue McKemmish	2/12/03	InterPARES meeting, Vancouver, B.C.
13.	Ross Harvey	2/17/03	Boston, MA

All interviews conducted by MVC (Smith, Marcum, Kenney, Jones, Frangakis, Sawyer, Bantin, Elkington, Dale, Waters, Harvey) or Sanett and Cloonan (Conway, Cox, Lesk, Besser, Turner, Lancashire, McKemmish).

November 21, 2003

Appendix 5: Mapping Table

Relationships between IP1 and IP2 Research Questions, IP2 Survey Topics and IP2 Research Questions, IP2 Key Informant Interview Questions and IP2 Research Questions

IP1 Research Questions (from Cloonan/Sanett study)	IP2 Research Questions (from Cloonan/Sanet t study)	IP2 Survey Topic & Questionnaire Sections (Project Focus)	IP2 Research Question Response	IP2 Key Informant Interview Questions (Big Picture Focus)	IP2 Research Question Response
What methods, procedures and rules of long-term preservation are in use or being developed?	4.1, 4.6, 4.7	B = Program & Policy	4.1, 4.6, 4.7	#1 Definition of Preservation (asked in Round 1 Survey)	n/a Det whether people's use of the term affected their approach to p.
What is the meaning of preservation?	n/a	C = Preservation Techniques	3.3(a), 3.3(b)	#2 Staffing, Education & Training (Survey F)	4.1 (a),4.1 (b), 4.8
Will current strategies for preservation ensure longevity and authenticity?	3.3(a), 3.3(b), 3.3(a)(b),4.1(a)(b), 4.6,4.7, 4.8	E = Cooperation	4.7, 4.8	#3 Preservation Planning (Survey N)	4.1, 4.6, 4.7, 4.8
How are costs for the preservation of electronic records derived?	3.3(b), 4.1(a)(b), 4.6,4.7,4.8	F = Staffing 7.4 Academic/professional backgrounds 7.5 7.5-6 Staff Training	4.8	#4 Prospective Future Users (survey K.12 &N)	4.1,4.6, 4.7, 4.8
Have effective cost models been developed?	n/a	G = Technical Questions	3.3(a)(b),4.1(a)(b), 4.6,4.7, 4.8		
		H = Cost Activities 9.8 Cost to Preserve Records 9.9 Cost Categories	3.3(b), 4.1(a)(b), 4.6,4.7,4.8	#5 Cost Activities (Survey H)	3.3(b), 4.1(a)(b), 4.6,4.7,4.8
		I = Records Organization/Provenance	3.3(a)(b), 4.1(a)(b),4.6	#6 Sources of Funding (Survey H & L)	4.6,4.7,4.8
		J = Description/Documentation of Preservation	3.3(a)(b), 4.1(a)(b),4.6	#7 Constraints (All Survey questions)	4.1 (a)(b), 4.6, 4.7, 4.8
		K = Access to Preserved Records	3.3 (a)(b), 4.1(a)(b), 4.6	#8 Greatest Concerns (All Survey questions)	
		L = Charges for Access	3.3(b), 4.1(a)(b), 4.6,4.7,4.8	#9 Current Projects (Any similarities to the projects described in the surveys?)	
		M = Reproduction & Copyright	3.3(a)(b), 4.1(a)(b)		
		N = Preservation Plans, Actions, Procedures and Policies	4.1(a)(b),4.6,4 .7,4.8		

Chart draft 7/31/03

Appendix 6: List of Changes Made from Round 1 to Round 2 of the "Preservation Strategies for Electronic Records" Questionnaire

Sections B & C: Respondents are asked for updates since round 1.

Section D: In round 1, this section was on Selection, an area eliminated in round 2. In the round 2 version of section D, respondents are asked to describe new cooperative partnerships.

Section E: In round 1, this section covered cooperative partnerships. In round 2, the section is on staffing. This is an expanded version of the staffing section which appeared as section F in round 1.

Section F: In round 2, this is the section on Technical Questions which was section G in round 1. The questions are the same in each version.

Section G: In round 2, this section covers Cost Activities; this was section H in round 1. The Cost questions have been expanded for round 2.

Section H: In round 2 this section is called, Records Organization and Provenance. In round 1 it was referred to as Preserving Records. The questions are the same.

Section I: In round 2, this is the section on Description/Documentation of Preservation Processes, which was section J in round 1. We have revised the questions for round 2 for purposes of clarification.

Section J: In round 2, this is the section on Access to Preserved Records which was section K in round 1. This section has been expanded for round 2.

Section K: In round 2 this section is called Preservation Plan, Actions, Procedures and Policies; it is an expansion of section N Preservation Policies from round 1.

The following sections from round 1 were eliminated:
L. Charges
M. Reproduction and Copyright

Appendix 7: Sources Cited by the Key Informants

The interviewees referred to the sources listed below during the course of the interviews. Their readings include professional, popular, and interdisciplinary works. Sometimes the interviewees cited just the author. For example, Anne Kenney talked about themes in the writings of Walt Crawford and Richard Cox mentioned the works of David Bearman without citing specific publications. Three electronic records research projects were also frequently mentioned in passing; these appear at the beginning of the list. The names of the interviewees who cited works appear in parentheses at the end of each citation.

With two exceptions, the key informants and interviewers did not cite themselves, but in group interviews some participants made reference to the works of others in the group. The general citation pattern that emerged is that interviewees cited either very recent works published since 2000, like Baker, Lessig, and Nichols/Smith, or else earlier works from the 1980s or the beginning of the 1990s that had influenced their thinking about the field. The one exception to this pattern was the *Preserving Digital Information* report edited by Don Waters and published in 1996. This report was mentioned by four of the interviewees, an indication of its continuing influence over the past seven years. Since four of the interviewees—Besser, Elkington, Lesk and Waters—were on the task force that produced the report one might expect the *Report* to have been brought up by them; but of the Task Force members, only Besser mentioned it. That other interviewees discussed it, shows the influence of this report beyond the immediate circle of the task force members.

The Three electronic records projects cited

University of Pittsburgh, "Functional Requirements for the Recordkeeping Systems," report published in July 2000. (Richard Cox and David Bearman, PI's) 1996.

Indiana University, "Electronic Records Project," Phase I, 1995-1997; Phase II, 2000-2002 (Philip Bantin, PI)

University of British Columbia, "The Preservation of the Integrity of Electronic Records," 1994-1997. (Luciana Duranti, PI, Terry Eastwood, co-PI, and Heather MacNeil, research assistant)

Works cited

Nicholson Baker, *Double Fold* (New York: Random House, 2001) (Cox; Kenney).

David Bearman's works. (Bantin, Cox)

Howard Besser, "Digital Preservation of Moving Image Material?" *The Moving Image* 1.2 (2001): 39-55. (Turner)

Michèle Valerie Cloonan, "Preservation Education in American Library Schools: Recounting the Ways," *Journal of Education for Library and Information Science* 31.3 (Winter 1991): 187-203. (Cox)

Paul Conway, "Overview: Rationale for Digitization and Preservation," in *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, MA: Northeast Document Conservation Center, 2000). (Conway; Conway's work in general was cited by Cox and Kenney)

Walt Crawford's works. (Kenney)

Pamela Darling, "To the Editor." *Conservation Administration News* 22 (July 1985): 3, and 20. (Frangakis)

Bruce W. Dearstyne, "What is the Use of Archives? A Challenge for the Profession." *American Archivist* 50 (Winter 1987): 76-87. (Conway)

Marilyn Deegan and Simon Tanner, *Digital Futures: Strategies for the Information Age.* (London: Library Association, 2002). (Elkington)

Jacques Derrida, translated by Eric Prenowitz. *Archive Fever: A Freudian Impression.* (Chicago: University of Chicago Pr., 1996). (Cox)

Marlan Green, et al. "Coming to TERM: Designing the Texas Email Repository Model," *D-Lib* 8.9 (September 2002): http://www.dlib.org/dlib/September 02. (Dale and Elkington together during their group interview)

Larry J. Hackman, "NHPRC Update," *History* News 34.1 (January 1979): 34-35. (Cox)

Miles Harvey, *The Island of Lost Maps: A True Story of Cartographic Crime.* (New York: Random House, 2000). (Cox)

Margaret Hedstrom, "Understanding Electronic Incunabula: A Framework for Research on Electronic Records," *The American Archivist* 54.3 (Summer 1991): 334-354. (Cox)

Larry Lessig, *Code and Other Laws of Cyberspace.* (New York: Basic Books, 1999).
_____. *The Future of Ideas.* (New York: Random House: 2001). (Waters)

Sue McKemmish and Frank Upward, *Archival Documents: Providing Accountability through Recordkeeping* (Melbourne: Ancora Press, 1993). (Harvey)

Stephen G. Nichols and Abby Smith, *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections.* (Washington, D.C.: The Council on Library and Information Resources, 2001). (Cox, Kenney)

Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action.* (New York: Cambridge University Press, 1990); and,

_____, and Robert Keohane. Eds. *Local Commons and Global Interdependence: Heterogeneity and Cooperation in Two Domains*, (London: Sage, 1995). (Waters)

Maxine K. Sitts, ed. *Handbook for Digital Projects: A Management Tool for Preservation and Access,* NEDCC, 2000. (Besser)

US-InterPARES Project, *Findings on the Preservation of Authentic Electronic Records: Final Report to the National Historical Publications and Records Commission.* (Conway, Harvey)

Don Waters and John Garrett, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.* (Washington, D.C.: Commission on Preservation and Access, 1996). (Besser, Conway, Harvey, Turner)

Citation of works by frequency

Waters	4
Conway	3
Baker	2
Bearman	2
Nichols/Smith	2
US InterPARES *Report*	2
Besser	1
Cloonan	1
Crawford	1
Darling	1
Dearstyne	1
Deegan/Tanner	1
Derrida	1
Green	1
Hackman	1
Harvey	1
Hedstrom	1
Lessig	1
McKemmishUpward	1
Ostrom	1
Sitts	1

Appendix 8: Key Challenges That Affect Digital Preservation

As identified by 18 key individuals interviewed
By Michèle V. Cloonan & Shelby Sanett,
August 2001 through February 2003

Technological Issues

- Obsolescence of hardware
- Computer viruses
- Computer crimes
- Media and software deterioration
- The debate over preservation strategies, e.g., migration, emulation
- Disappearing information (both file formats and entire infrastructures)
- Interrelation problems (where are the boundaries of the work?)

- Maintaining relationships or links between digital content and digital metadata
- Metadata (definition; the need for crosswalks)
- Intellectual property issues ("the muddy waters of rights management")

- Short length of life of media coupled with our lack of recognition that it is more than the media itself that needs to be preserved
- How to preserve through time the key characteristics of records

Strategic Issues

- Identifying what needs to be preserved
- Custodial problems (who is responsible for saving it?)
- Dependence on global context
- Economic issues (where does the responsibility lie for support?)
- All technology depends on technical support, which requires stable funding
- Archivists/librarians must keep making the case for preservation
- Lack of preservation awareness on the part of the public
- No clearly defined methods or best practices
- Lack of standards
- Lack of a research agenda
- Ignorance; e.g. educating administrators/users to understand that digitization is not preservation
- Management of the process
- Mindset (educating professionals to be pro-active)

Archival Issues

- Selection and appraisal (what to keep?)
- Deciding what is of value
- Conceptual issues (what is an archive? what is a digital archive?)
- Evidence and accountability
- Authenticity and reliability (trustworthiness)