Digital Preservation Needs and Requirements in RLG Member Institutions

by Margaret Hedstrom and Sheon Montgomery
December 1998

A study commissioned by the

Research Libraries Group

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Libraries, archives, and museums play a critical role in organizing, preserving, and providing access to the cultural and historical resources of society. Digital technologies are used increasingly for information production, distribution, and storage. The institutions that have traditionally assumed responsibility for preserving information face technical, organizational, resource, and legal challenges in taking on the preservation of digital holdings.

The Research Libraries Group (RLG) is an international consortium of universities and colleges, national libraries, archives, historical societies, museums, independent research collections, and public libraries. Its mission is to "improve access to information that supports research and learning"—through collaborative activities and services that include organizing and preserving as well as sharing information resources. In early 1998, RLG funded a study by Dr. Margaret Hedstrom and Ms. Sheon Montgomery of the status of digital archiving in its member institutions. The primary purpose was to assess where guidance, education and training, storage, and digital preservation services are required, in order to develop the kind of training, mechanisms for resource sharing, and services that meet members' needs.

Fifty-four institutions participated in the survey and 15 in targeted, supplementary interviews. Of these, two-thirds already assume responsibility for preserving material in digital form; by 2001, 98% expected to be preserving both acquired or "born-digital" items and materials they have converted to digital form. Only half of the institutions with digital preservation responsibilities have policies that govern acquisition, conversion, storage, refreshing, and/or migration of digital materials. Less than half of the institutions with digital holdings refresh them by copying to new media or migrate these materials to current formats. Most institutions that do refresh or migrate digital materials carry out these activities ad hoc or in conjunction with system upgrades, rather than as an integral part of a digital preservation program. Almost half of member institutions with digital holdings also lack the capacity to mount, read, or access files on some of the storage media they hold. Three-quarters of them believe irreplaceable information will be lost if digital preservation issues are not resolved. Collection managers view technology obsolescence as the greatest threat, followed by insufficient resources and insufficient planning. The need for digital preservation expertise is high: asked to rate staff as expert, intermediate, or novice, only eight of the 54 institutions considered their staff at the expert level.

RLG's members identify a variety of measures to respond to anticipated growth in digital acquisitions and conversion. All of the institutions that currently have written policies for preserving digital materials, as well as 33 of 36 that do not, anticipate developing new policies in the next three years. They plan on staff training, hiring additional, expert staff, and using consultants. There is consensus that coordinated strategies and shared resources are essential to achieving broad solutions and enhancing local efforts; but there is

anxiety over the present lack of sufficient standards and concrete guidelines for collections and service suppliers.

This report examined one component of an evolving infrastructure for long-term preservation of digital information: the responsibilities of archives, libraries, museums, and other repositories for preserving and providing access to valuable, digital resources. Although there is a gap between current models for digital preservation and the status of digital preservation in many institutions, most member institutions are seeking guidance on ways to close this gap. The members participating in this study look to RLG to make available concrete standards, guidelines, and training that will enable institutions at various stages in their digital preservation programs to work with confidence; that are flexible enough to evolve as technology and sources of information evolve; and that can be used to help ensure successful, quality services from third-party vendors. RLG is also seen as experienced in and committed to international coordination and to integration of archival, museum, and special collections into the mainstream of digital preservation activities. And members rightly expect that RLG will use consortial leverage on their behalf to identify and make real the standards and supporting services needed in a digital world.

INTRODUCTION

Libraries and archives play a critical role in organizing, preserving, and providing access to the cultural and historical resources of society. In the relatively stable world of printed, hand-written, and mechanically reproduced information, repositories managed to preserve a rich array of scholarly communications, documentary evidence, and useful information for specialized scholars and for the general public. The introduction of digital technologies into the processes of production, distribution, and storage of information challenges the capacity of libraries, archives, museums, and other cultural institutions to carry out their responsibilities for preservation. This problem has been the focus of numerous reports designed to raise awareness of digital preservation issues and to propose general strategies for addressing them.[1]

The general outline of digital preservation challenges is well established. Digital materials are especially vulnerable to loss and destruction because they are stored on fragile magnetic and optical media that deteriorate rapidly and that can fail suddenly from exposure to heat, humidity, airborne contaminants, or faulty reading and writing devices. Even if the media are preserved intact, digital materials become unreadable if the playback devices necessary to retrieve information from the media become obsolete or if the software that translates digital information from machine- to human-readable form is no longer available. Libraries, archives, and other repositories that traditionally have assumed responsibility for preserving information face technical, legal, and organizational challenges in responding to the new demands of digital preservation. Repositories need access to technical resources—both information systems that support digital preservation and the technical expertise to use these technologies effectively. Institutions also face legal obstacles in fulfilling their mandates to preserve valuable information when copyright or licensing agreements prohibit duplication or local storage of digital information. Institutions can take action on their own, but there is a strong consensus that coordinated strategies and shared resources are essential to achieving broader solutions to digital preservation and enhancing the success of local efforts.[2]

Survey Purpose and Methodology

In early 1998, the Research Libraries Group (RLG) funded this study of digital archiving needs and requirements in RLG member institutions. RLG is an international consortium of close to 160 members, including universities and colleges, national libraries, archives, historical societies, museums, independent research collections, and public libraries. The purpose of the study was twofold: (1) to gather baseline data on the nature and extent of digital preservation problems in member institutions and the status of their digital preservation programs, and (2) to identify needs and requirements of member institutions in meeting their responsibilities for preserving digital information. The authors and RLG were particularly interested in learning whether digital preservation is a common concern across libraries, archives, museums, and special collections, or whether

this problem is still limited to large institutions that were early adopters of digital technologies. They also wanted to gather data about the policies and practices that are being used to preserve digital materials, in order to determine the extent to which successful models and prescriptive guidelines are known and are being replicated. Finally, they hoped to gain a deeper understanding of obstacles to digital preservation in member institutions.

This study builds on the May 1996 *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, which was co-sponsored by the Commission on Preservation and Access and RLG, and it complements a number of other follow-on studies.[3] That report recommended development of a deep infrastructure capable of supporting a distributed system of digital archives. The report drew attention to the need for a sufficient number of trusted organizations capable of storing, migrating, and providing access to digital archives as essential elements of a digital archiving system. Research for this 1998 report on preservation practices and needs in RLG member institutions addresses a number of questions related to developing an infrastructure for digital preservation. Identifying areas of common concern among member institutions suggests areas where development of common preservation strategies and methodologies might prove beneficial; where RLG and other consortia can play a useful role in developing standards, best practices, and services for digital preservation; and where there is potential for development of services and products to facilitate digital preservation.

Methodology

The research was conducted by Margaret Hedstrom, Associate Professor at the School of Information, University of Michigan, and Sheon Montgomery, Graduate Student Research Assistant. The methodology used for this research was a combination of a survey distributed to all RLG member institutions and follow-up interviews with one or more administrators at 15 institutions. This combination of data collection methods was used to gather and analyze data on the nature and types of digital collections in member institutions, the policies and programs in place for digital preservation, digital preservation needs and requirements, and future plans for developing digital preservation programs. The interviews with administrators provided additional qualitative insights into the problems they face and their perceptions of various strategies for digital preservation.

Scope and Conduct of the Survey

Libraries, archives, museums, and other repositories acquire digital materials through several different channels, which include purchasing digital information from publishers or distributors, legal deposit or transfer, donations, and licensing access to online databases. Many institutions also are creating digital information through programs or projects that convert print, photographic, and manuscript material to digital form. Libraries and archives do not necessarily assume responsibility for long-term preservation of all these materials. Some print-to-digital conversion projects exist primarily to improve access to materials, and some institutions

consider the originals or a microfilm copy to be the preferable version for long-term preservation. Online databases that are updated regularly may only be useful in their current version. Licensing agreements and copyright restrictions governing online digital resources and hand-held digital products might prohibit institutions from maintaining copies of digital works for preservation purposes.

For the purposes of this project, the survey focused on digital materials for which an RLG member institution assumes long-term preservation responsibility. The materials within the scope of the study include: (1) information created originally in digital form and acquired by a repository through purchase, transfer, or legal deposit; and (2) digital materials that are created when repositories convert print, photographic, and manuscript materials to digital form, provided that the institutions considered it their responsibility to preserve the information.

The term "digital preservation" refers to both preservation of materials that are created originally in digital form and never exist in print or analog form (also called "born-digital" and "electronic records") and the use of imaging technology to create digital surrogates of analog materials for access and preservation purposes. While this broad use of the term digital preservation can cause confusion, data on both aspects of digital preservation were analyzed. Digital materials, regardless of whether they are created initially in digital form or converted to digital form, are threatened by technology obsolescence and physical deterioration.

The survey was designed to gain an understanding of member institutions' current holdings of digital materials and to gather data on present and near future needs related to preservation of digital materials. The survey questions were grouped into the categories of:

- digital materials policy,
- digital holdings,
- storage methods and formats,
- digital knowledge/training, and
- future needs for services and training.
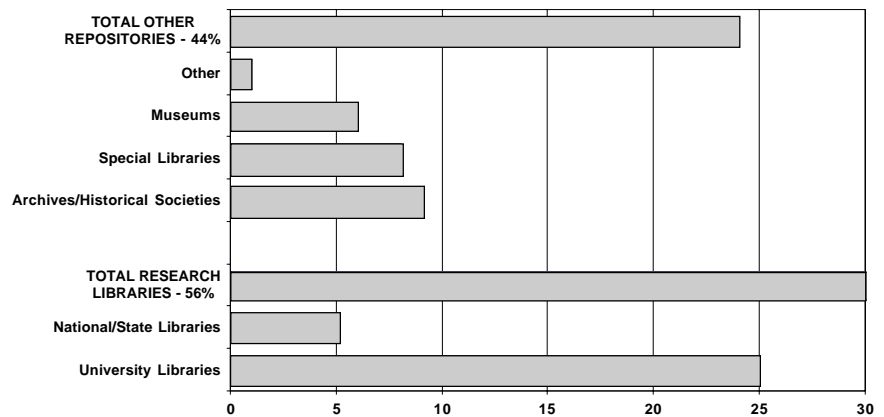
Survey Administration

A survey, accompanied by an introductory letter from Robin Dale, RLG Program Officer for Member Initiatives, was distributed in late April 1998 to the designated member representative in all 158 member institutions (see Appendix A). E-mail was the primary distribution method, but copies were distributed in print form via fax or express mail to institutions without e-mail access. A test survey consisting of the scope, definitions of terminology, and 25 questions had been distributed by RLG in March 1998. Based on feedback from eight individuals, the survey was revised and simplified to its final form of 18 questions, a statement about the scope of the survey, and definition of six key terms used in the survey. The survey was designed for ease of completion on a computer screen and reply via

e-mail, with provision for returning the survey via fax or postal service. Although a May 6 deadline was specified, all surveys received by the end of August 1998 were included in the data analysis.

Response Rate and Survey Population

A total of 54 completed surveys were received between April and August 1998 for a response rate of 34%. Forty-two surveys were returned from members in the US and 12 from members in Australia (1), Canada (2), The Netherlands (1), Switzerland (1), and the UK (7); see Appendix B.

Figure 1: Distribution by Institution Type of 54 Survey Respondents

TOTAL OTHER REPOSITORIES - 44%	
Other	
Museums	
Special Libraries	
Archives/Historical Societies	
TOTAL RESEARCH LIBRARIES - 56%	
National/State Libraries	
University Libraries	

Based on an informal classification, the responses are distributed by institution type as shown in Figure 1: National or State Libraries (5), University Libraries (25), Archives/Historical Societies (9), Special Libraries (8), Museums (6), Other (1).

Interviews

Interviews were conducted to gain additional insights that would supplement the quantitative data from the surveys. The interview script consisted of six open-ended questions designed to encourage spontaneous and reflective responses. Three questions focused on the current situation within the institution regarding:

- the impact of digital preservation demands on the interviewees and their institutions,
- experience(s) with outsourcing services, and
- availability and effectiveness of technical support within the institution.

One question asked administrators for feedback on the major issues that their institution will be facing with digital materials during the next four years; another focused on their insights

4

regarding consortial arrangements for preservation of digital materials. The final question was a wish-list scenario, situating respondents "outside the box" to voice concerns about a particular problem regarding preservation of digital materials that they would like to see "magically solved."

RLG provided a list of 25 member representatives who were targeted for interviews. An e-mail message was sent to all, requesting their participation in individual thirty-minute interviews. Representatives of 16 institutions responded, and it was possible to schedule interviews with 14 of these. A test interview was also done, and as it was determined that no changes to the script were necessary, this was included in the final compilation for a total of 15 interviews. All interviews were conducted over the telephone, with the interviewer typing responses immediately into Microsoft Word. Fourteen interviews were completed between April 21 and May 7; the final interview was completed on June 15. Three interviews were conducted as conferences with two or three representatives from different departments in one institution. Interview length was determined by the interviewee, and ranged from approximately 15 minutes to one-and-a-half hours, with the average being about 25-30 minutes. At the end of each interview the interviewee was asked to return the institution's survey so that complete data would be represented in the analysis. The non-respondents in this interviewed group were the only institutions to which follow-up appeals were made to return surveys. Of the 15 institutions that provided interviews, 12 institutions are represented in the survey data.

Data Analysis

The survey data was compiled in a database in Microsoft Access. This database was then exported into a Microsoft Excel spreadsheet, and all statistical comparisons were made using formulas in a series of spreadsheets.

Owing to their qualitative nature and unstructured format, the interviews were more challenging to record. While complete interview responses were available in their entirety, for ease in comparison the major components of each response were distilled into abbreviated phrases (maximum 75 characters), coded by institutional identifier and question number, and arranged in an Excel spreadsheet. This produced a brief document that could be quickly scanned, with the responses arranged in columns by question number, and in rows by institution. Based on analysis, 10 major themes were then identified, and a second spreadsheet was created in which the coded, abbreviated phrases were matched with the themes they represented. This again produced a document in which responses could be easily scanned by theme, as well as by institution identifier and question number.

The data gathered and analyzed in this research permits drawing several conclusions about the current state of digital preservation in RLG member institutions and about their perceptions of needs and requirements for digital preservation. There are three aspects of the research that may limit broad generalizations from the results.

Survey Population

The sample was drawn from RLG member institutions to provide RLG with analysis of conditions and needs of its member institutions. The approach has the advantage of gathering data from a variety of institutions with digital preservation responsibilities, including research libraries, archives and historical societies, special research collections, and museums. However, there may be systematic differences between RLG member institutions and similar institutions that are not RLG members; and consequently, findings from this survey may not apply to all libraries, archives, museums, and special collections with digital preservation responsibilities.

Size of Sample and Response Bias

Although 34% is a reasonably good response rate for a survey of this sort, there are two potential problems with the number of respondents and the nature of the response rate. The total of 54 responses allows detecting patterns and drawing certain conclusions across the entire data set. However, when one breaks the data into smaller categories by type of institution, location, whether or not the institutions have holdings or policies, etc., the small number of cases in some categories makes it difficult to draw statistically significant conclusions. Given this limitation, many of the findings are descriptive of the entire population, and trends within categories are suggested where the data supports such conclusions.

The potential bias in response is toward institutions with some digital materials in their holdings for which they assume preservation responsibility. Although there is no way of comparing the characteristics of non-respondents with institutions that responded, it is possible that institutions which consider digital preservation an important problem because they have digital materials in their holdings were more motivated to respond to the survey. This cautionary note is reinforced by anecdotal information gathered from respondents while they were completing the surveys and through informal feedback received after the surveys were completed. Several institutions reported difficulty completing the survey because there was no single individual or source of information for answering all of the questions on the survey instrument. Compiling all of this information and answering the questionnaire thoroughly required a significant amount of time on the part of member institutions, suggesting that a high level of motivation was necessary for member institutions to respond.

Shared Concepts and Terminology

Digital preservation is a multi-faceted problem that is viewed differently by different institutions and different professionals. The broad umbrella of digital preservation includes preserving both materials created and acquired in digital form and digital files generated when institutions convert analog materials to digital form. The survey instrument took special measures to explain the scope of the survey and to define the technical terminology used. While acknowledging that there are different interpretations of the terminology used, it requested the respondents to refer to and use the definitions provided. Nevertheless, one cannot be certain that all respondents shared exactly the same concepts and terminology in all of their responses.
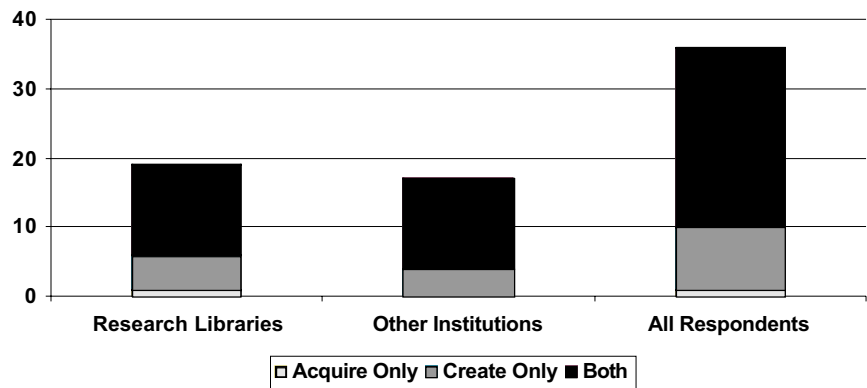
**Digital Preservation
Responsibilities**

Digital preservation is not an isolated problem affecting only large libraries and archives. Two-thirds (36) of the institutions in the survey have some digital materials in their holdings for which the institution assumes preservation responsibility. In fact, the 30 research libraries, which tend to be larger than the 24 archives, museums, and special collections, are not quite as likely to hold digital materials as the other types of institutions. About two-thirds of the research libraries assume responsibility for preserving digital materials, whereas 71% of the other types of repositories hold digital materials. A higher percentage of institutions outside the US have digital preservation responsibilities—83% as compared to 62% of the institutions in the US.

Most repositories with digital holdings both acquire material in digital form and create digital files through conversion (see Figure 2). Two-thirds of the institutions with digital holdings both acquire and create digital materials. One-quarter of the institutions are

**Figure 2: Acquisition and Creation of Digital Holdings by Institution Type in 36 Institutions**

responsible for preserving only information that they create through conversion. Two institutions acquire information in digital form, but do not create digital files through conversion.

In the survey overall, almost half of the member institutions acquire digital materials and nearly two-thirds are involved in conversion activities. Through both acquisition and conversion activities, RLG member institutions are taking on responsibility for preserving unique digital materials that appear not be duplicated elsewhere. Twenty-eight institutions reported that irreplaceable information would be lost if the digital materials in their holdings are not adequately preserved for future use.

The digital holdings that member institutions are responsible for preserving vary in size, age, and the types of formats used for storage. Taken together, RLG member institutions that responded to the survey hold a minimum of 275,000 digital files and a total of at least 4.1 terabytes of information. These estimates are stated as minimum values because complete information on the quantity of digital materials was not available. Many institutions do not know how much digital material they hold and those that do use different measures to keep track of the quantity of digital materials in their holdings. The survey asked institutions to report the number of digital files, the number of volumes (reels of tape, CDs, etc.), and the total storage volume (in MB, GB, etc.). Only 26 of the 36 institutions with digital materials were able to estimate the quantity of material using any one of these measures. Holdings range in quantity from less than 200 files at one institution to 96,600 files in another repository. Three repositories account for almost 75% of the total number of files and 90% of the total volume of 4.1 terabytes. In the 19 institutions that estimated the quantity of their holdings in megabytes or gigabytes, storage volumes range from a low of 500 megabytes in one institution to a high of 2 terabytes (see Figure 3).
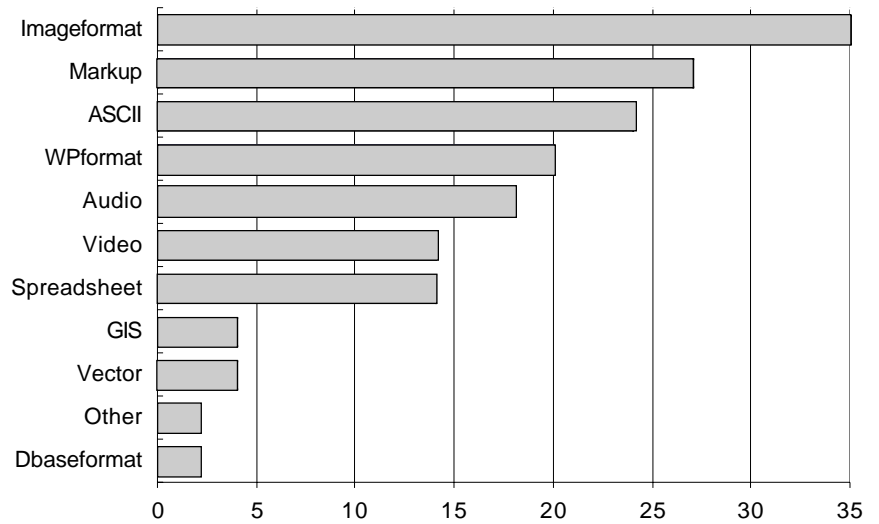
Figure 3: Estimated Volume of Digital Holdings in Gigabytes in 19 Institutions
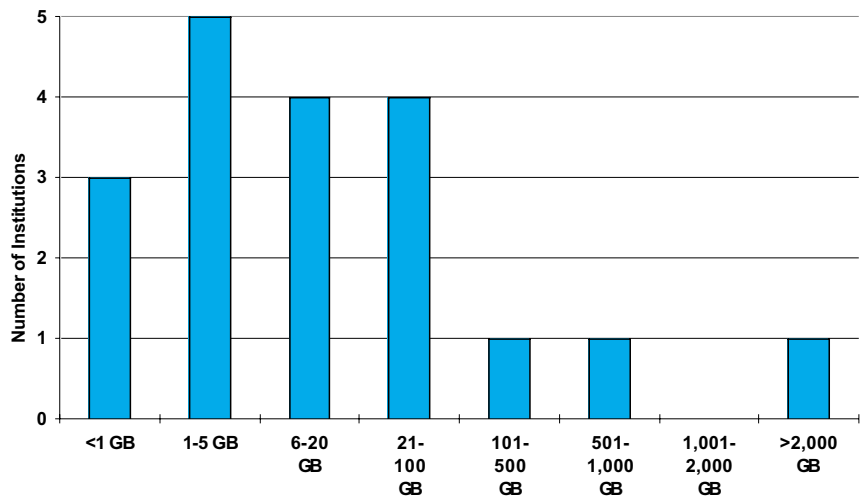
**1,000 Megabytes (MB) = 1 Gigabyte (GB)   1,000 GB = 1 Terabyte (TB)**

Many file formats and storage media are present in the digital holdings of these institutions (see Figure 4). Among the 36 institutions with digital holdings, the most common formats are image files (97.2%), text files with mark-up (75.0%), and ASCII files (66.7%). More than one-third of the institutions had digital information in at least one of the following formats: word processing files (55.6%), audio (50.0 %), video (38.9%), and spreadsheets (38.9%), while fewer than a third of the institutions maintain digital information in geographic information systems, vector graphics, databases and other miscellaneous formats.

Figure 4: Distribution by Storage Format of Acquired Digital Holdings in 36 Institutions

| Format	Value
Imageformat	35
Markup	27
ASCII	24
WPformat	20
Audio	18
Video	14
Spreadsheet	14
GIS	4
Vector	4
Other	2
Dbaseformat	2

When institutions create digital information through conversion, they also create archival master files in a wide variety of formats for preservation (see Figure 5). The most common formats for archival master files of converted images are TIFF (80.6%); text files with mark-up (55.6%); JPEG (38.9%); and PDF, GIF, and ASCII (30.6% each).

Figure 5: Distribution by Storage Format of Digital Archival Masters Created via Conversion in 34 Institutions

Format	Value
TIFF	28
Text	19
JPEG	13
PDF	10
GIF	10
ASCII	10
Other	7
EPS	5
MPEG	4
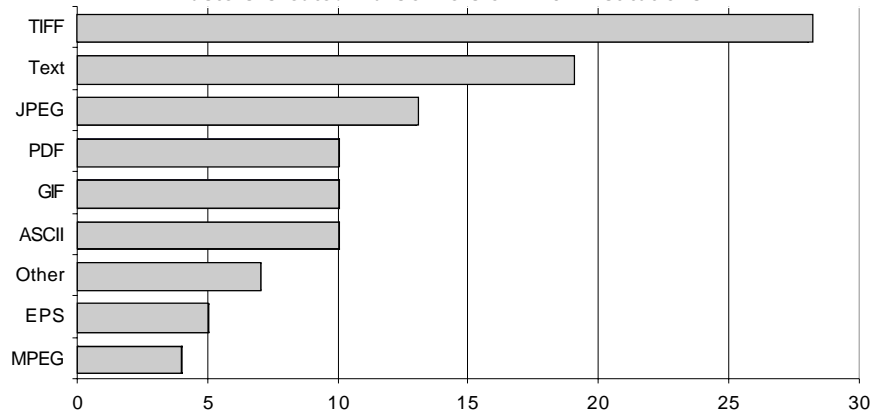
The survey shows that at least 24 different storage formats are present across the 54 responding institutions.[4] Many institutions are maintaining digital information in several different formats. As shown in Figure 6, 24 of the 36 institutions with digital holdings maintain digital information in at least six different formats, and 13 of these institutions have 10 or more different formats.

10

**Figure 6: Number of Formats of Digital Holdings
in 36 Institutions**
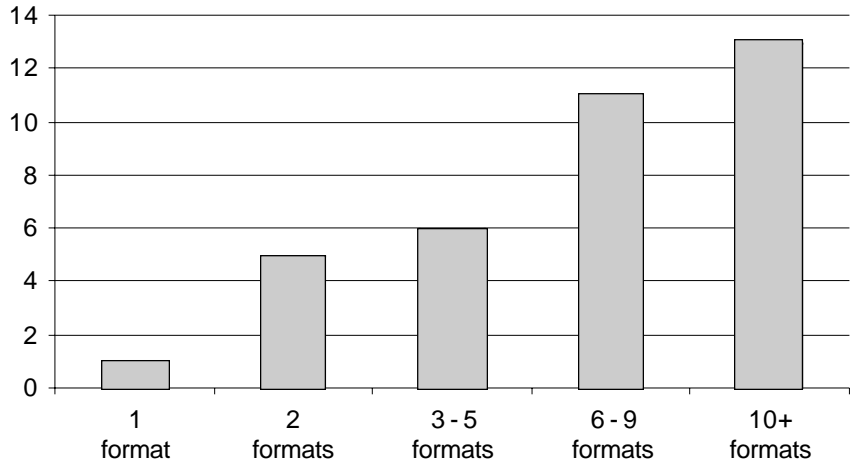
Most of the institutions with digital holdings maintain relatively recent material, with some notable exceptions. The oldest digital materials reported were written to their current storage medium in the late 1970s, but the vast majority were generated in the last one to three years (see Figure 7). Of 33 institutions reporting on the age of their oldest digital materials, 15 have materials created since 1995; only seven have materials created prior to 1990. Institutions with recent acquisitions and those that started conversion projects only in the last few years are just beginning to confront problems of long-term management of digital information.

**Figure 7: Distribution by Date of Oldest Digital
Materials in 33 Institutions**

Creation Date of Oldest Digital Materials

11
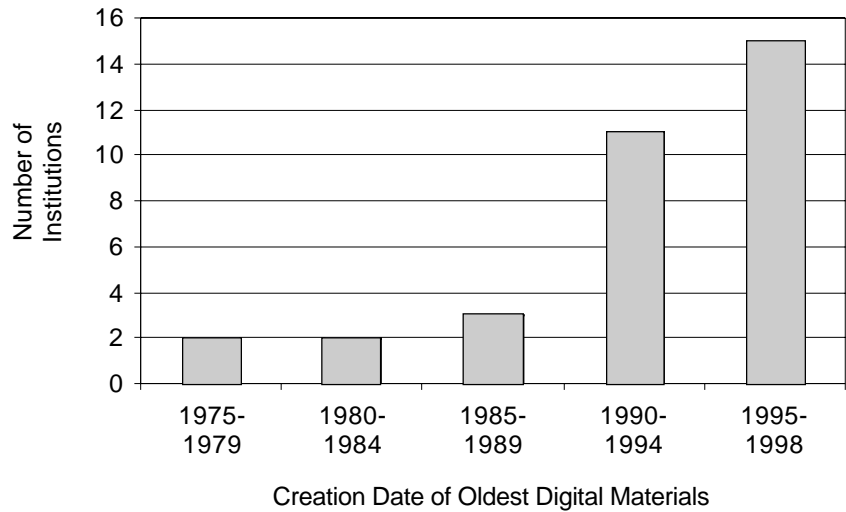
In spite of the relatively young age of digital collections in most institutions, 15 institutions reported that they lack the operational and/or technical capacity to mount, read, and access some digital materials in their holdings. The most common storage media for which institutions lack access capability are floppy disks (five institutions) and open reel nine-track tape (three institutions); there are also scattered problems with CD-ROMs, magneto-optical disks, DDS DAT tape, 3480 cartridges, and various audio and video formats.

Summary | This research on digital preservation responsibilities and the nature and extent of digital holdings provides concrete evidence about digital preservation problems facing libraries, archives, museums, and other repositories. Digital preservation is a widespread problem in RLG member institutions. Typically, institutions with digital preservation responsibilities both acquire materials in digital form and create digital files through conversion. While a few institutions have large collections of digital materials, most are modest in size but growing rapidly. Most member institutions began acquiring or creating digital materials during the last one to five years. Nevertheless, 15 of the 36 institutions with digital holdings cannot access some of their materials because they lack the operational or technical capacity to mount, read, or access files stored on some of the storage media in their holdings. Three-fourths of the institutions with digital holdings report that irreplaceable information will be lost if the digital materials in their holdings are not adequately preserved for future use.

Digital Preservation Policies and Practices

Digital preservation policies and practices are not well developed in member institutions. Two-thirds of the institutions do not have written policies for digital preservation. One common reason that institutions appear not to develop digital preservation policies is that they have not yet assumed responsibility for preserving materials in digital form. However, taking responsibility for digital preservation does not necessarily mean that institutions use policies to govern their digital preservation activities. Only half of the institutions with digital materials in their holdings have written digital preservation policies (see Figure 8).

Figure 8: Digital Preservation Policies for Digital Materials Held in 36 Institutions
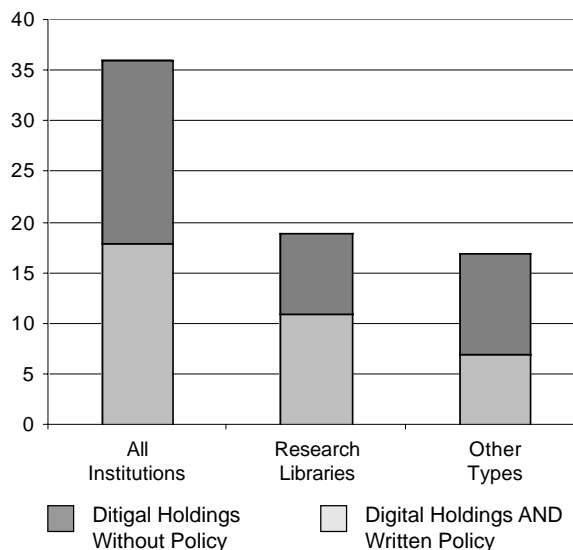
Fourteen institutions with policies considered their policy adequate, two institutions said it met the institution's needs poorly, one said well, and one was not sure. As shown in Figure 9, most of the policies govern the acquisition of materials in digital form, but less than half of the policies govern storage of digital information, conversion of materials from print to digital form, refreshing, and migration. Only five institutions have policies that cover acquisition, storage, migration, and refreshing. Of the 34 institutions that are generating information in digital form through conversion, only seven have policies governing conversion.

Figure 9: Comprehensiveness of Written Policy for Managing Digital Holdings in 18 Institutions
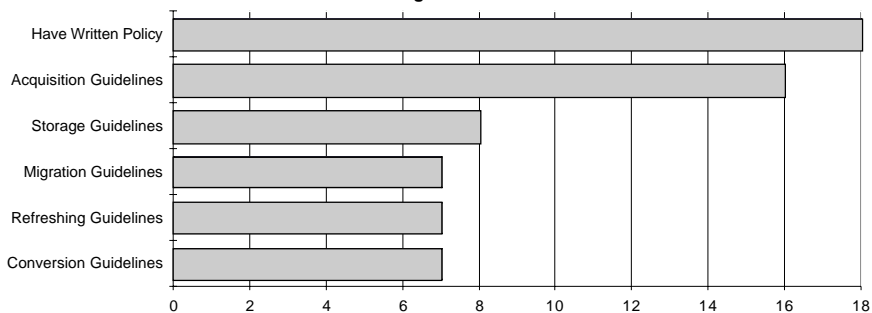
This data and additional information gathered through interviews allows speculation about the lack of written policies. Most institutions that have not yet acquired materials in digital form or created digital materials through conversion have not yet experienced the need for written policies. The interviews suggest that among those institutions with digital preservation responsibilities, the lack of good models for digital preservation and confusion about the most appropriate methods and approaches are major obstacles to developing effective policies and practices. Several administrators who were interviewed believe that policy development is important, but they do not yet see stable standards or effective models that can be used as the basis for institutional policy. Some expressed concern over competing strategies and the wide variety of approaches being advocated or followed by different institutions. At the same time, there is little information available on the costs or the effectiveness of different strategies. For some institutions, the perceived lack of a consensus in the community serves as a disincentive to policy development.
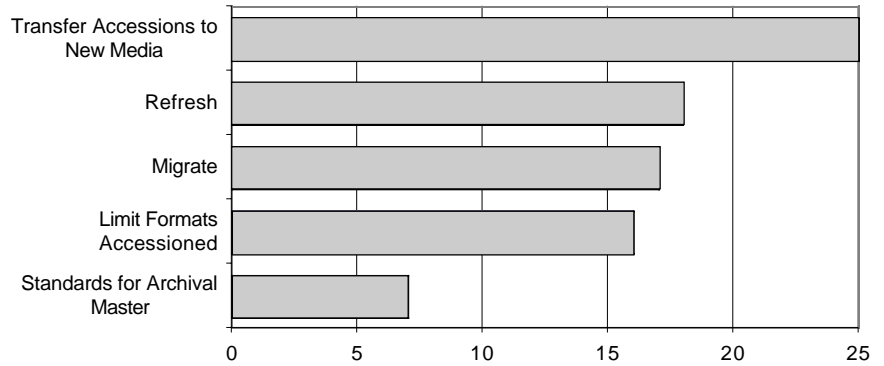
Preservation Practices

Effective digital preservation requires life-cycle management of digital information from the point of creation through storage, migration, and providing access on a continuing basis. Only 13 institutions reported that they have established methods in place for digital preservation. Figure 10 indicates the types and extent of methods in use. Some of the institutions in the survey have adopted practices that limit the formats of digital information that they must manage. Two institutions limit the acceptable formats to flat files while others accept several different formats such as PDF, TIFF, SGML, and word processing formats. Only seven of the 34 institutions that create materials through conversion have adopted a standard format or formats for preservation master files, and there are a wide variety of formats present across the institutions.

The institutions in the survey use a wide range of methods for storing the digital information in their holdings. It appears that a majority of institutions store some materials on the media in which they are received and transfer some materials to another digital medium, most commonly magnetic tape or cassette (68%), a hard drive (64%), or CD-ROM (52%). One-quarter of these institutions transfer some files to open reel magnetic tape. The use of rewritable magneto-optical disk and WORM optical disks is uncommon. One institution reported using a robotic tape storage system,[5] while another prints its digital information on paper for preservation. Four institutions contract with a third party for storage. While there are several different storage methods in use, most institutions are avoiding proprietary and highly vulnerable media, such as magneto-optical and WORM disks.

Most institutions do not have established methods for digital preservation beyond transferring some material to new storage media. Eighteen institutions "refresh" digital materials, which the survey defined as copying digital materials in their original digital format from old to new media. A wide variety of practices are used for refreshing digital information. Nine institutions have refreshing

**Figure 10: Preservation Methods Used for Digital
Holdings in 36 Institutions**

Method	Value
Transfer Accessions to New Media	25
Refresh	18
Migrate	17
Limit Formats Accessioned	16
Standards for Archival Master	7

cycles that range from as often as daily or weekly as part of the routine operations of a hierarchical storage management system to as infrequently as once every 10 years. The most common refreshing cycle is in the range of once every three to six years. One institution cleans and rewinds magnetic media on a yearly rotation and copies those files that exhibit access problems or physical deterioration. Two institutions reported that they refreshed digital materials only as a by-product of a larger system upgrade, while two others reported that the problem of refreshing had not arisen yet because of the newness of digital activities. One institution reported that refreshing services were provided by centralized information technology services and that the details of their practices were not available.

Seventeen institutions also reported that they migrate digital materials. The survey defined migration as the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation.[6] Migration practices also vary among institutions and they do not appear to be well integrated into regular digital preservation processes. Rather, migration has mostly occurred on an ad hoc or project basis. Most institutions that described their migration methods indicated that they performed migration as necessary to ensure continuing access or as a by-product of a larger system upgrade. One institution reported that they had migrated digital images from a Xerox proprietary format (XDOD) to TIFF 5.0/6.0 and another institution was launching a similar project. When migration takes place, it appears to occur in response to impending technological obsolescence or in conjunction with larger system upgrades.

Summary

The policies and practices for digital preservation are underdeveloped in member institutions, especially given the increasing prevalence of digital materials in their holdings. Those institutions with the largest digital collections also appear to have the most developed policies and practices, and those without holdings have not yet experienced the need to address digital preservation issues. Most institutions have not adopted policies that limit the variety of digital formats and storage media and few have adopted standards for the

master files of materials that they generate through conversion. As a consequence, there are a significant number of institutions that maintain digital materials in many different formats. Many of these files are relatively new and the institutions have not yet confronted problems of technology obsolescence or systems upgrades. When migration is necessary or when the systems are upgraded, these institutions may face a complex and expensive task of converting files from many different formats, including proprietary formats. While more than half of the institutions with digital holdings refresh and/or migrate digital materials, these activities are not well integrated into digital preservation programs and tend to occur on an ad hoc basis or in conjunction with system upgrades.

Digital Preservation Staffing and Expertise

Lack of staff expertise is a common problem both in institutions with digital preservation responsibilities and in institutions that have not yet assumed responsibility for digital materials. As shown in Figure 11, only eight institutions ranked the highest level of digital preservation knowledge on their staff as expert, while 21 ranked the highest level of in-house knowledge at the intermediate level. The largest group of institutions (23) reported that the level of staff expertise was at the novice level, and two reported that they had no expertise available.

Figure 11: Levels of Digital Expertise Within 54 Institutions
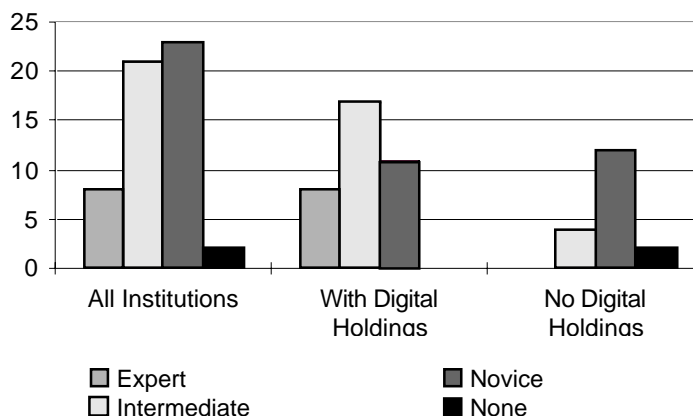

There is a relationship between having responsibility for preserving digital materials and levels of expertise. All institutions with experts available and 80% of the institutions with intermediate-level expertise have digital materials holdings in their holdings. This is

16

true of only half of the institutions with novice-level expertise available and neither of those with no expertise. It is not clear from the data, however, whether taking responsibility for preservation encourages institutions to develop the requisite expertise, or whether institutions are reluctant to assume responsibility for digital preservation until at least a modicum of expertise is available in their institutions.

The telephone interviews gathered data about additional sources of support and about the experience of institutions that had used third-party services for some aspect of digital preservation. Some institutions get technical support from centralized institutional computing services such as the campuswide computing services in universities. Interviewees were asked about the role of institutional computing services in meeting digital preservation needs. About two-thirds reported that they used institutional computing services, most often for backup, telecommunications, setting up systems, and as a source of expertise. Several of the interviewees mentioned that although they have access to centralized computing services, these services do not meet the needs of the repository. One interviewee pointed out that the computing services organization is responsible for meeting the institution's business needs and not the special requirements of digital preservation. Another mentioned the "intermediate view" of computing services, but a lack of awareness of long-term preservation. Several respondents reported that, as organizations, the centralized computing services are also small or understaffed and lack technical expertise, while two people mentioned a "lack of imagination" and a conservative orientation among centralized computing staff. It appears that centralized computing services are potential sources of support for some digital preservation programs, and several people reported that they had initiated or established relationships with these organizations. Nevertheless, centralized computing organizations usually cannot meet all of the digital preservation needs of repositories.

Twenty institutions also use external sources such as consultants or contracts with third-party vendors to gain access to expertise in digital preservation or to outsource certain digital preservation functions. Interestingly, those institutions with the highest levels of staff expertise are most likely to acquire expertise from outside sources (see Figure 12): three-fourths of the institutions with in-house digital experts do so, as well as almost half of those with intermediate-level in-house expertise and one-fifth of those with novice-level in-house staff. Neither of the institutions with no in-house expertise available use outside sources for digital expertise.

**Figure 12: Correlation Between In-House Digital Expertise
and Use of Outside Sources in 54 Institutions**
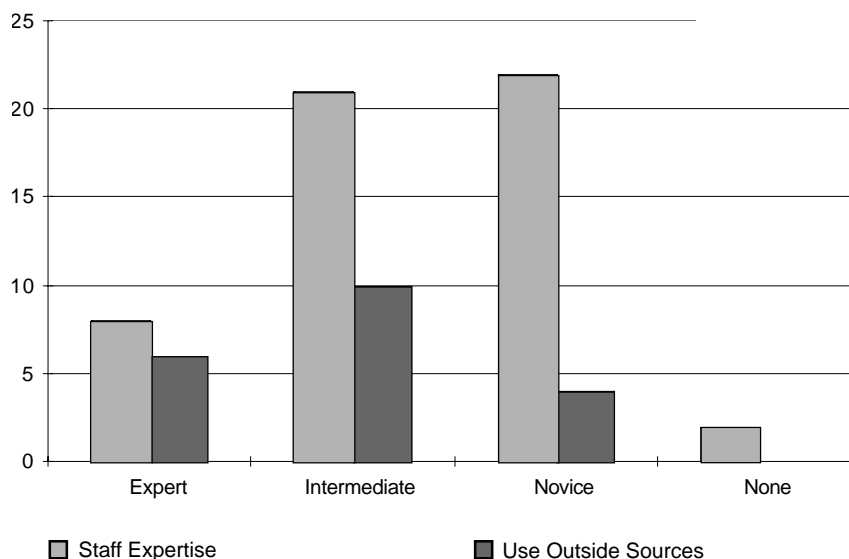
Interviews also yielded some insights into the experiences of member institutions that had outsourced preservation activities or contracted with third parties. A few institutions had outsourced digital conversion and preservation microfilming, but no institutions have experience with outsourcing digital preservation activities. Several interviewees expressed an interest in outsourcing conversion and one mentioned services related to access. The interview research suggests some of the concerns about outsourcing. A few interviewees mentioned the need for quality services and for agreements between the institution and the vendor about acceptable quality. Others mentioned that the third-party services have to be flexible, affordable, and beneficial to the institution. Several respondents also discussed the potential for a consortium of institutions to share leads and experiences with vendors and to develop quality services without redundancy.

Summary The lack of expertise in digital preservation appears to be a significant obstacle to developing digital preservation programs. The programs with the highest levels of staff expertise also tend to have significant holdings of digital materials. Those institutions with expert staff also are more likely to take advantage of outside experts by hiring consultants with expertise to supplement that available on their staffs. Although this research did not collect quantitative data specifically on the expertise available from centralized institutional computing services, the information gathered through interviews suggests that centralized computing services provide needed expertise and services only in exceptional cases. Some institutions have developed or are starting to form effective relationships with centralized computing services, but computing services often lack the staff, specific knowledge of digital preservation requirements, and vision to meet the digital preservation needs of many institutions. Finally, institutions have limited experience using third-party service providers for digital preservation. This may be due

primarily to the lack of service providers in this area, but concerns over quality control and cost are also deterrents to outsourcing digital preservation services.

Needs and Requirements

The survey and interviews captured information about administrators' perceptions of digital preservation problems, their anticipated changes in digital preservation responsibilities in the next three years, and their sense of the most important issues facing their institutions. This research also gathered data on plans for increasing the institutions' capacity for digital preservation and respondents' perspectives on various types of services that consortia and/or third-party providers might develop to assist member institutions to implement sound digital preservation programs.

Problems and Threats to Digital Preservation

Respondents were asked to rank the significance of four issues as threats to digital preservation: technology obsolescence, insufficient resources, insufficient planning, and physical condition of materials (see Figure 13). Member institutions ranked technological obsolescence as the greatest threat to loss of digital materials, followed closely by insufficient resources and an insufficient policy or plan for digital preservation. Only one institution ranked physical condition of the materials as the most serious threat. The lack of resources for digital preservation is ranked as the greatest threat by institutions that have policies in place for digital preservation and by the institutions that have digital materials in their holdings. Although the size of the sample is too small to reach statistically significant conclusions, this suggests that as institutions assume responsibility for preserving digital materials and as they develop policies, they then confront the resource requirements in concrete terms.

Figure 13: Threats to Digital Holdings as Ranked by 54 Institutions

Legend:
- Greatest Threat
- Greater
- Lesser
- Least Threat
- Average Ranking

19
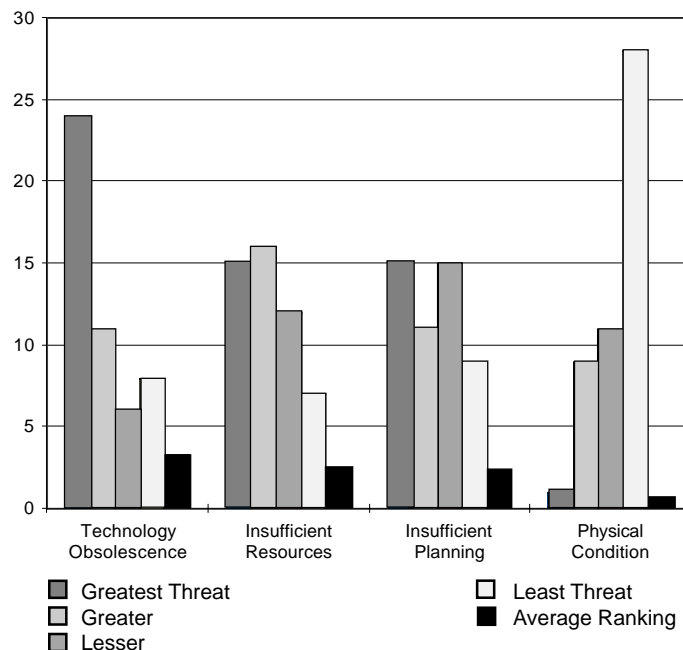
The interviews with managers of digital collections provide additional insights into how this problem is perceived and which strategies institutions are using to address it. When asked about the impact of digital preservation on their institutions, the interviewees mentioned changes in the way information management and preservation are organized and carried out, additional resource requirements, changing relationships with other professionals and specialists, and the need to hire new staff with digital preservation expertise. Several digital collection managers mentioned the lack of standards, planning, infrastructure, and models for digital preservation as significant problems for their institutions. Several administrators in archival repositories expressed concern about the changing nature of institutional memory, where paper records are disappearing but the repositories are not prepared to handle electronic records.

Another common theme is the strain that digital preservation places on resources, including staff, equipment, and funds. Several administrators mentioned the need to hire staff with technical expertise and a few had plans in place to do so. Cooperative efforts with experts in technology or with other individuals responsible for digital preservation were also mentioned as strategies to share expertise and learn from others. Several administrators mentioned the increasingly complex relationships between libraries or archives and information technology and service providers. As responsibility for digital preservation becomes spread across several different divisions there is a need for more coordination. Two interviewees also raised lack of awareness and support from top administrators as a problem for dealing with digital preservation in a concrete manner. Some interviewees felt that administrators paid lip service to digital preservation issues but were not willing to support serious institutional efforts to address the problem. Finally, a few interviewees mentioned the positive impact of digital initiatives on their institutions, especially in the way that conversion of materials to digital form offers new means to deliver information to users and to preserve the originals.

The respondents had mixed reactions about the impact of digital preservation on them personally and professionally. They were fairly evenly divided between those who found digital preservation interesting and stimulating and who had made large personal investments to keep up with the issues, and those who were concerned about the challenges, the absence of clear guidance, and the need for greater expertise. Many of the interviewees said that digital preservation was forcing them to re-examine traditional practices, change the way they administer their departments, develop more interdepartmental relationships, and learn new skills.

Anticipated Changes in Responsibility

As Figure 14 illustrates, member institutions appear to take digital preservation seriously. Almost all of the institutions in the survey anticipate an increase during the next three years in the quantity of digital materials that they are responsible for preserving.

Fifty-two institutions expect an increase in digital materials from conversion projects, while 47 institutions expect digital preservation

**Figure 14: Anticipated Increase in Digital Holdings by 2001
in 54 Institutions**
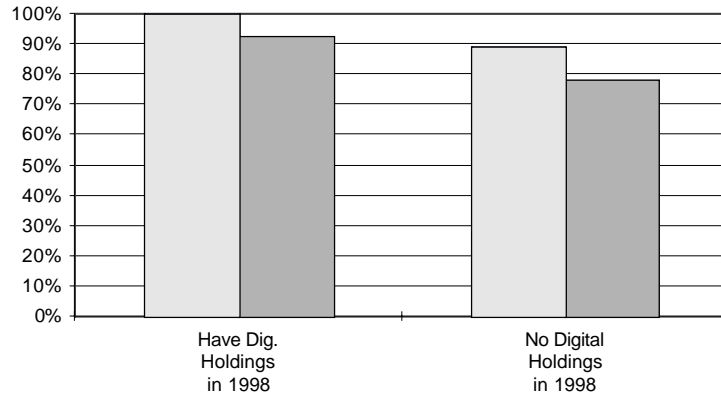
Note: 98% of surveyed institutions (53 of 54) anticipate having digital materials in their
holdings by 2001, as compared to 67% with digital holdings in 1998 (36 of 54).

☐ via Creation   ☐ via Acquisition

responsibilities to grow as a result of new acquisitions or accessions of electronic records.

In the interviews with administrators, several mentioned new acquisitions or the increasing size of the collections as a major issue that they would have to confront in the next four years. An archivist in a state archives anticipates that with a change in administration, records in electronic form will begin to arrive in the archives for the first time. The institution needs to develop information systems immediately that can support cataloging, access, and preservation of this material. In a different state archives, the archivist considered the need for more staff and funding to deal with the growing number of digital records as a major issue facing the institution. In this case, developing a fairly technical system to describe, preserve, and provide access to digital materials is considered a priority so that it is possible to make access more self-service. The director of a university library that may be designated as a national repository for digital materials considers the problem very serious because the library will have to be able to take responsibility for digital preservation if required to do so through legislation.

When asked about the major digital preservation issues facing institutions in the next four years, the digital collection managers who were interviewed mentioned a wide variety of issues along several common themes. A few respondents stressed the importance of raising the level of awareness about digital preservation problems, especially among top administrators. Some collection managers expressed concerns that institutions would sidestep the issue of long-term preservation by viewing digital technologies as a way to provide access to materials, but that at some point this approach would fail, leaving institutions with a preservation crisis. This presents a challenge of convincing administration to see digital preservation as a priority and to take some immediate steps to avoid even greater problems in a few years.

Several interviewees discussed human resource implications, including the need for more staff to carry out labor-intensive digital work, the demand for staff with new and different skills, the challenges of "upskilling" and retraining, and the need to redefine job descriptions and skill requirements.

A third theme relates to the absence of standards, clear models and best practices, especially in the areas of migration and frequent updating of hardware and software. Several digital collections managers stressed the importance of working together as a community to drive the development of needed products, services, and international standards so that vendors will begin to create products that meet digital preservation needs. The interviewees also mentioned the need to address preservation in the context of materials that institutions license, but do not own; better coordination among the various parties involved in digital preservation; and the development of tools for appraisal and risk assessment.

Proposed Actions | Most member institutions are planning to take concrete measures to build up their capacity for digital preservation. Almost all of the institutions (51 of 54) anticipate developing new policies for preserving digital materials in the next three years. This includes all of the institutions that currently have written policies as well as 33 of 36 that do not currently have a policy.

Institutions also anticipate using a variety of methods to increase the level of available staff expertise with digital preservation. Thirty-one institutions hope to hire people with digital preservation knowledge or expertise. As Figure 15 shows, institutions plan to use multiple sources for training and staff development, ranging from training provided by professional organizations (42 institutions) to hiring consultants (20 institutions.)

Figure 15: Preferred Methods for Increasing Staff Digital Expertise in 54 Institutions
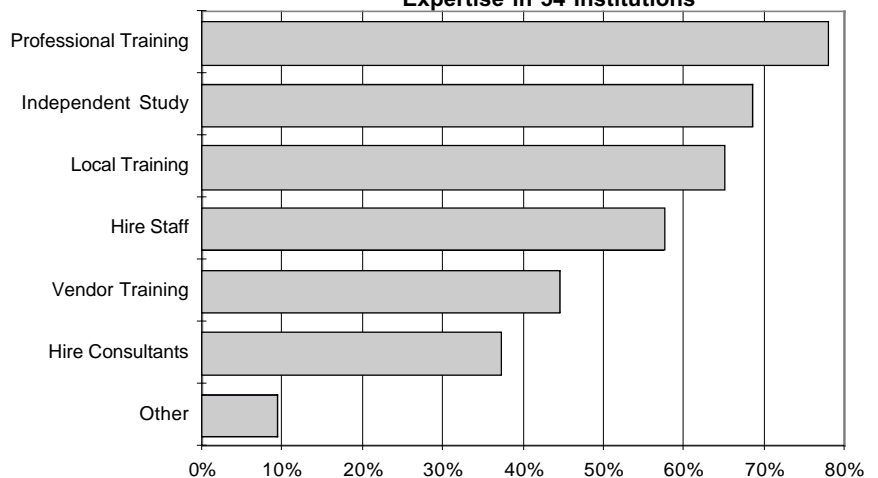
Institutions also expressed an interest in using a variety of digital archiving services if they were made available at a reasonable cost. The broad interest in model policies, standards and best practices, training, and specific storage, maintenance and conversion services suggest useful roles for institutions, consortia such as RLG (see Figure 16), and commercial service providers (see Figure 17).
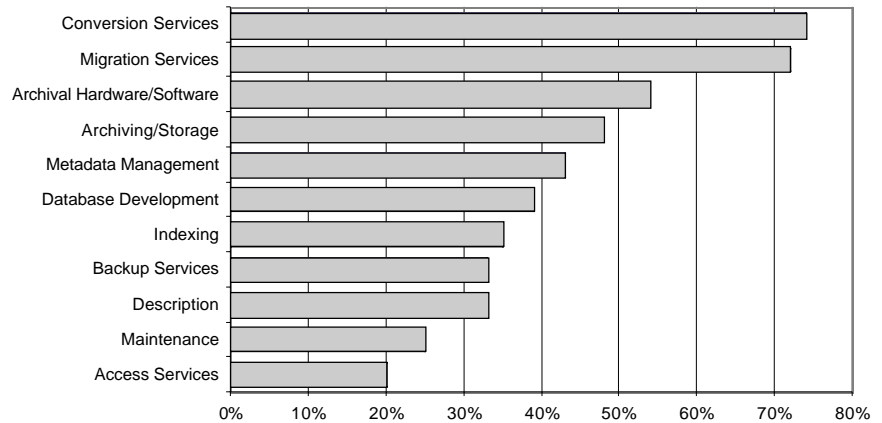
In our interviews we also asked about attitudes toward the possibility of institutions pooling resources into a consortium or network for digital preservation, and how each interviewee would see his or her institution fitting into such a network. In general, the interviewees supported cooperative efforts, but they differed in their views of how such cooperative efforts should be organized and exactly which activities could be done best through a consortium. The strongest support was expressed for having a consortium that could provide leadership and coordination in best practices and standards, serve as a source of technical information and expertise, provide training, and bargain collectively with publishers and vendors for more affordable services. The response was mixed to the idea of having a consortium assume responsibility for storage, preservation, and access to digital materials. While a few interviewees said they would be relieved if a trusted organization were available for storage and management of digital materials, many interviewees expressed concerns about losing control over their material.

Figure 16: Digital Archiving Services Preferred from Consortia in 54 Institutions
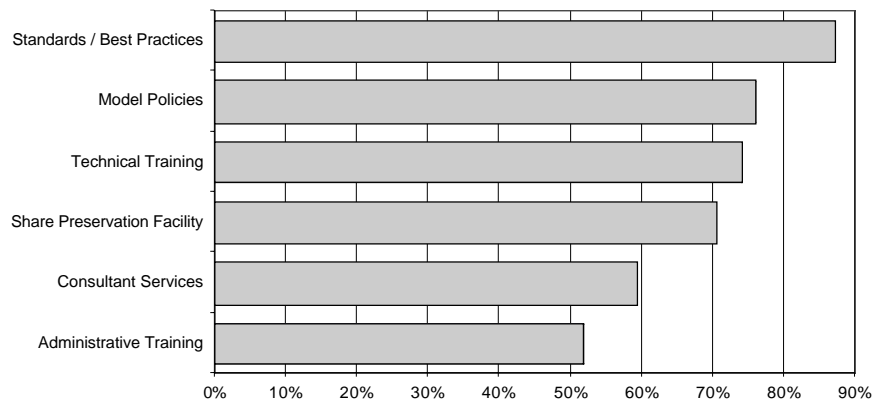
Interviewees favored the idea of pooling resources to eliminate redundant efforts and reduce preservation costs, but they expressed several concerns. One manager pointed out that outsourcing and pooling of resources is popular politically, but it is difficult to achieve. This represents a major change from the way institutions have worked traditionally. Some of the keys to developing successful cooperative efforts appear to be a clear definition of the role and responsibilities of any consortium-based service and a proven ability to provide a quality service at a lower cost than the institutions can achieve on their own.

Figure 17: Digital Archiving Services Desired from Third-Party Vendors in 54 Institutions

Service	Percent
Conversion Services	~74%
Migration Services	~72%
Archival Hardware/Software	~54%
Archiving/Storage	~48%
Metadata Management	~43%
Database Development	~39%
Indexing	~35%
Backup Services	~33%
Description	~33%
Maintenance	~25%
Access Services	~20%

This research also compiled a "wish list" of issues that digital collection managers would like to see solved. When asked to choose "one concern or problem to be immediately and magically solved," most interviewees had trouble limiting their response to a single issue. Interviewees provided a lengthy wish list of issues to solve, but some issues appear repeatedly among the top concerns. One request is for a clear, coherent set of guidelines and standards to follow. Some respondents were frustrated by the multitude of competing strategies and the varied and confusing voices addressing digital preservation. Several respondents considered technology obsolescence the most important issue to resolve, but they proposed different priorities for resolving this problem. One interviewee wanted to avoid migration by using standards as a way to ensure continuing access to materials. Another stressed the need for cost models and process models to deal with migration and obsolescence issues. Two interviewees mentioned fast, reliable and stable storage media.

There was also considerable interest in dealing with the economics of and funding for digital preservation. First, there is an interest in better data on the costs of digital preservation and the need for cost and funding models which in turn might be used to increase institutional understanding and commitments to digital preservation. This is often coupled with a desire for more resources and for more reasonably priced services.

Finally, several respondents mentioned legal and organizational issues, including the changing nature of ownership of information, the role of publishers and their responsibility for preserving digital products, and the issues of access, licensing, and copyright. Several interviewees also stressed the need for a long-term vision oriented toward the future to avoid rapid obsolescence of guidelines, resources, and technology.

Summary Member institutions expect their digital preservation responsibilities to increase during the next three to five years. During this time period, all but one institution in the survey anticipate some involvement with digital materials through acquisition, the creation of digital materials, or both. Institutions that already have responsibility for digital preservation expect that their digital holdings will grow during this time period. Many institutions that only recently started digital initiatives will confront the need to convert materials from older formats to current formats for the first time.

The institutions that responded to the survey are planning a number of actions to prepare for increased digital preservation responsibilities. Many institutions plan to hire additional staff with expertise in digital preservation, provide training to current staff, and hire consultants to help with the development of digital preservation programs.

Member institutions are seeking leadership in the development of standards and best practices, guidance on model policies and practices, and various types of training from consortia such as RLG. The respondents also expressed an interest in a variety of services from third-party vendors, especially conversion services, migration services, hardware and software to meet archival needs, and archiving or preservation storage services. Acceptance of third-party services is contingent on reliability and a reasonable cost.

RECOMMENDATIONS

This research supports a series of recommendations and potential next steps for RLG and other consortia, for member institutions, and for commercial service providers. The following list of recommendations is by no means exhaustive, and it is derived from the types of data gathered through this survey. Issues such as changes in legislation, the development of specific media, systems interoperability, metadata standards, and the need for more research are beyond the scope of this survey. The authors' work confirms that legislation, standards, and research are important elements of a digital preservation infrastructure.

Recommendations for RLG

1. *Compile and distribute a set of guidelines, standards, and best practices for digital preservation.*

Member institutions would like concrete guidelines, standards, and best practices for digital preservation. Many of the respondents to the survey considered RLG a good source for such guidelines based on RLG's previous contributions to standards for preservation, microfilming, and archival description. Member institutions would like guidelines that are authoritative and specific enough so that staff who are still developing skills in digital preservation can follow and implement them. At the same time, guidelines or best practices have to take into account wide variations in the scale of digital preservation problems, the formats of digital material, and organizational arrangements for digital preservation. Guidelines and best practices should be flexible enough to evolve as member institutions increase their digital holdings and as digital preservation methodologies improve.

2. *Provide leadership and coordination in emerging standards and practices for digital preservation.*

It would be unrealistic to expect RLG to assume an exclusive leadership role in emerging standards and practices for digital preservation; the issues are too complex. Several other organizations, including the Council on Library and Information Resources, the Association of Research Libraries, the Coalition for Networked Information, and the Consortium of University and Research Libraries, are also providing leadership. But there are important areas where leadership is needed and RLG is the logical organization to assume a leadership role. Two of these are in coordinating international digital preservation initiatives and in integrating archives, museums, special research collections, and small libraries into the mainstream of digital preservation activities.

The research suggests that while there are slight differences in some of the attributes of US members and those outside the US, member institutions share a common set of needs and requirements, and many institutions recognize the need for global solutions. RLG, as an international organization with 24% of its members outside the United States, is well situated to provide leadership internationally

that will help reduce redundant efforts at developing solutions, accelerate the adoption of best practices, and enhance access to digital materials via interconnected global networks.

RLG also has experience working with archives, museums, special collections repositories, and smaller organizations that are not always served by models and services designed for large research institutions. This study demonstrates that special repositories are as likely to face digital preservation problems as larger institutions and that they are seeking leadership and coordinated initiatives to help them solve digital preservation problems.

3. *Investigate further RLG's potential role as a provider of specific services and shared resources.*

Training, coordination, and advice are the areas where the largest percentage of member institutions looks to RLG for assistance. Several members expressed an interest in having RLG assume the role of a direct service provider in storage and maintenance of digital files, but a significant number of institutions raised concerns about loss of control over their digital materials and about quality control in a largely untested area of third-party service provision. As an alternative to providing services directly, RLG might begin by representing members' interests to third-party providers and bargaining on their behalf for more reasonably priced products and services.

Recommendations for Member Institutions

1. *Develop means to coordinate digital preservation activities within institutions.*

Digital preservation activities are dispersed in many institutions among different administrative units and different types of professional staff. The challenges that many institutions faced in responding to the survey provide evidence of the lack of coordination of these activities. Digital preservation is likely to remain a distributed responsibility because of the range of expertise needed to address this problem. Nevertheless, coordinating mechanisms within institutions would enhance utilization of existing resources, foster more consistent policies and practices, draw attention to this issue at higher levels of administration, and in some cases help achieve economies of scale.

2. *Develop institutional policies for acquisition, conversion, storage, and maintenance of digital materials.*

Comprehensive policies for acquisition, conversion, storage, and maintenance are lacking in many institutions. As a consequence, many institutions are acquiring digital materials or generating digital materials through conversion without a strategy for how to preserve them. Limiting the variety of storage formats that are acceptable for acquisitions and adopting standard formats for the archival master files of converted images can reduce future storage and maintenance problems. Assigning responsibility for storage, maintenance, and migration can help institutions avoid crises and minimize the impact of unanticipated changes in hardware and

software. Explicit policies on the scope of digital collections and on digital conversion goals are useful for planning within institutions and for building a framework of shared responsibilities across institutions.

Recommendation for Service Providers

Develop trusted services and tools to support digital preservation in critical areas of need.

Research identified several areas where institutions are seeking affordable services to support digital preservation. Conversion services, consulting on digital preservation, and reliable, affordable storage are desired by many member institutions. In addition to an interest in more affordable services for conversion of print, photographic, and manuscript materials to digital form, it appears that there will be a large demand for conversion of existing digital materials from obsolete formats to emerging open or standard formats. It appears that many institutions will need tools for converting materials that are in proprietary word processing formats and for maintaining various versions of TIFF images.

CONCLUSION

The Task Force on Archiving of Digital Information called for a deep infrastructure of institutions, services, technologies, and qualified personnel capable of supporting a distributed system of digital archives. This report examined one component of this infrastructure in detail: archives, libraries, museums, and other repositories that have been instrumental in preserving and providing access to scholarly communications, documentary heritage, and other cultural resources in traditional formats. It is clear that many of these institutions are beginning to add digital preservation to their array of preservation responsibilities, although most are taking only the first steps in this direction. Since 1995, the number of institutions that acquire digital materials and that convert traditional formats to digital form has more than doubled. Only one respondent does not anticipate taking responsibility for digital preservation within the next three years. By 2001, 98% of the institutions included in this survey will assume responsibility for preserving some digital information.

There is a gap between current models for digital preservation and the status of digital preservation in many institutions. Institutions with large digital collections and more years of experience generally have policies in place that govern acquisition, storage, refreshing, and migration of digital materials. But the majority of institutions have not developed digital preservation policies or established methods to preserve digital information.

Several factors help explain the slow pace of development of digital preservation programs. For many institutions, digital preservation is a new challenge and they are just beginning to confront the policy, technological, and human resource implications of digital preservation. Institutions are seeking better methods and more affordable services to tackle the problems posed by technology obsolescence. Insufficient resources and inadequate planning for digital preservation also are considered major obstacles to digital preservation, and only a few institutions have experts in digital preservation on their staff or available through consultation arrangements. Finally, the absence of a clear consensus about effective strategies and methods for digital preservation and the paucity of data on the resource implications of various proposed strategies serve as deterrents to concrete actions by institutions.

A critical mass of institutions has taken responsibility for preserving digital materials. In spite of considerable variety in the types of institutions and in the size, age, and formats of their digital collections, there are also ample opportunities for developing common solutions and for sharing resources. Leadership from RLG and other organizations in developing and promoting standards and best practices for digital preservation, as well as reliable and affordable services from third-party providers, are essential components of the evolving infrastructure for preserving distributed digital collections.

NOTES

[1] Committee on the Records of Government, *Report*, Washington, D.C.: The Committee (1985). Commission on Preservation and Access and The Research Libraries Group, Inc., *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, Washington, D.C.: CPA and RLG, May 1996. Association of Research Libraries, *Transforming Libraries: Issues and Innovations in Preserving Digital Information*, Washington, D.C.: Association of Research Libraries, December 1997. Arts and Humanities Data Service, *Strategic Framework for Creating and Preserving Digital Collections*, July 1998, available November 1998 at http://ahds.ac.uk/manage/framework.htm. National Research Council, *Preserving Scientific Data of Our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources*, Washington, D.C.: National Academy Press, 1995.

[2] *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, pp. 37-41; *Transforming Libraries: Issues and Innovations in Preserving Digital Information*, p. 7.

[3] *Transforming Libraries: Issues and Innovations in Preserving Digital Information*. Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation,* Washington, D.C.: Council on Library and Information Resources, available January 1999 at http://www.clir.org/. *Strategic Framework for Creating and Preserving Digital Collections.* Consortium of College and Research Libraries (CURL), the *CEDARS Project: CURL Exemplars in Digital Archiving*, available December 1998 at http://www.leeds.ac.uk/cedars/; David Haynes and David Streatfield, "A National Coordinating Body for Digital Archiving?" *Ariadne*, 15, available at http://www.ariadne.ac.uk/issue15/digital/.

[4] Twenty-four different file formats were reported by the respondents through a combination of checking off formats listed in the survey instrument and listing additional formats that had not been identified. This figure understates the variety of formats present because some general categories, such as "word processing" formats, did not ask respondents to indicate which version of specific formats were present.

[5] A robotic tape storage system is usually a form of hierarchical storage management system that uses a combination of online, near-line, and deep storage.

[6] *Preserving Digital Information*, p. 6.

April 1998 Survey Letter

Digital Preservation Needs Assessment

Introduction

In 1996, the Task Force on Archiving Digital Information, co-sponsored by RLG and the Commission on Preservation and Access, released its analysis of major digital archiving issues. As one of its follow-up activities, RLG is sponsoring a survey of the digital archiving needs of member institutions. This survey is directed by Margaret Hedstrom, Associate Professor in the School of Information at the University of Michigan and is being conducted by Graduate Student Research Assistant, Sheon Montgomery. Hedstrom and Montgomery will analyze the data from this survey and prepare a report for RLG's Digital Archiving Working Group. The report will be available to RLG member institutions later this year.

We would appreciate your assistance with this survey. Please send your response via e-mail to: shmontgo@umich.edu no later than MAY 6, 1998.

Purpose of this survey:

The purpose of this survey is to assess the digital preservation needs of RLG member institutions, including their needs for guidance, education and training, storage, and preservation services. Results of the survey will be used to develop training, mechanisms for resource sharing, and services that meet the needs of member institutions.

Scope of the survey:

The survey is designed to gather data about the formats and the condition of those digital materials in your institution's library and archival holdings and for which your institution assumes responsibility for preservation. This includes materials that your institution acquires in digital form, as well as materials that your institution creates originally in digital form, or through digital conversion.

NOTE: Please exclude those digital materials to which your institution provides access but does not assume responsibility for long-term preservation.

This survey is being distributed to each RLG Institutional Member Representative. If you are unable to respond, please forward this survey to the most appropriate person(s) in your institution. If there are multiple repositories at your institution which hold digital materials, please indicate which repositories are included in the survey.

Thank you for your time and cooperation.

Robin Dale, RLG Member Services Officer

Terminology used in this Survey:

digital conversion projects = scanning materials (e.g., books, in this Survey newspapers, manuscripts, graphic/pictorial/photographic works) to create digital files

digital materials = information sources in digital form, including converted materials and electronic records

electronic records = materials originally created in digital form

migration = periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to subsequent generation

refreshing = copying digital files in their original digital format from old to new media

preservation = long-term storage, maintenance and migration of digital materials

NOTE: We realize that there are different interpretations of the terminology provided above. Please refer to these definitions when responding to the survey to increase consistency of the results.

Digital Materials Policy:
1. Does your institution currently have any written policies for managing digital materials?
    ( ) No, skip to #2   ( ) Yes

1a. If YES, does policy provide guidelines for: (check all that apply)
    ( ) acquiring materials in digital form
    ( ) converting materials from print to digital form
    ( ) storage
    ( ) refreshing
    ( ) migration

1b. If YES, how well does this policy meet your institution's current needs?
    ( ) Well   ( ) Adequately   ( ) Poorly

Current Holdings:
2. Does your institution have any digital materials in its holdings for which it assumes responsibility for their preservation?
    ( ) No, skip to #12   ( ) Yes

3. Does your institution accept or acquire (through donation, legal deposit, or purchase) electronic records for which it assumes preservation responsibility?
    ( ) No   ( ) Yes

3a. If YES, do you accept electronic records in any format, or only in specified formats?
    ( ) Any   ( ) Specified Only—please indicate:

4. Does your institution currently create digital materials as a result of digital conversion projects or by any other conversion methods?
    ( ) No   ( ) Yes

Storage Methods and Formats:

5. Which of the following formats are present in the digital holdings for which your institution assumes preservation responsibility? (check all that apply)
   - ( ) Flat ASCII files
   - ( ) Text files with markup (e.g., SGML, HTML, XML, etc.)
   - ( ) Wordprocessing format (e.g., MS Word)
   - ( ) Database format (e.g., Access, FoxPro)
   - ( ) Spreadsheet format (e.g., Excel)
   - ( ) Image format (e.g., TIFF, GIF, etc.)
   - ( ) Geographic Information System (e.g., ARC INFO)
   - ( ) Audio
   - ( ) Video/Moving Images
   - ( ) Vector graphics (e.g. CAD files)
   - ( ) Other—please specify:

6. What method(s) do you use to store electronic records (i.e., those materials received in digital form)? (check all that apply)
   - ( ) Store as received
   - ( ) Transfer to other digital storage medium - please indicate:
   - ( ) Hard drive
   - ( ) Magnetic tape (open reel)
   - ( ) Magnetic tape (cassette or cartridge)
   - ( ) CD-ROM
   - ( ) Optical Disc (Rewritable)
   - ( ) WORM Optical Disk (Write-once-read-many)
   - ( ) Other—please specify:
   - ( ) Contract with 3rd party for storage
   - ( ) Other method—please specify:

7. If your institution creates digital materials, what format(s) do you use for preservation purposes (i.e., archival master files)? (check all that apply)
   - ( ) Don't know
   - ( ) ASCII
   - ( ) Text files with mark-up (HTML, SGML, XML, etc.)
   - ( ) EPS (Encapsulated Postscript)
   - ( ) TIFF
   - ( ) GIF (Graphics Interchange Format)
   - ( ) JPEG
   - ( ) MPEG
   - ( ) PDF
   - ( ) PICT
   - ( ) WMF (Windows Metafile)
   - ( ) Image Pac
   - ( ) Other, please specify:

8. In what year were the oldest digital materials in your holdings written to their current storage medium?

8a. What is their current storage medium and format?

8b. Are there any digital materials in your holdings for which you lack the operational and/or technical capacity to mount, read, and access?
   - ( ) No ( ) Yes

9. Does your institution have an established method for preserving digital materials?
    ( ) No   ( ) Yes

9a. Does your institution refresh digital materials?
    ( ) No   ( ) Yes—please describe frequency/method:

9b. Does your institution migrate digital materials?
    ( ) No   ( ) Yes—please describe frequency/method:

10. Can your institution determine or estimate the quantity of digital materials for which you currently have preservation responsibility?
    ( ) No   ( ) Yes: (Please report any data that is readily available):

    _____ approximate number of unique files

    _____ approximate number of volumes (reels of tape, optical disks, etc.)

    _____ total storage volume (in MB, GB, etc.)

10a. For what percentage of your digital holdings do back-up copies exist?
     ( )%

11. If the digital materials at your institution are not adequately preserved for future use, will irreplaceable information be lost?
    ( ) No   ( ) Yes

Digital Knowledge/Training:
12. What is the highest level of knowledge available in-house for digital preservation activities?
    ( ) Expert   ( ) Intermediate   ( ) Novice   ( ) None

13. Does your institution currently utilize outside sources of expertise for preservation of digital materials (e.g., consultants, contracts)?
    ( ) No   ( ) Yes

Future Needs:
14. Do you anticipate developing new policies for digital materials within 3 years?
    ( ) No   ( ) Yes

15. Over the next 3 years, do you expect an increase in the quantity of your digital holdings (i.e., for which your institution will assume long-term preservation responsibility) due to …

    …acquisitions/accessions in electronic form?
    ( ) No   ( ) Yes

    …conversion projects (i.e., from print to digital)?
    ( ) No   ( ) Yes

16. How would you rank the following factors as threats to the loss of digital materials at your institution within the next 3 years? 1 = greatest threat, 4 = smallest threat
    ( ) Physical condition
    ( ) Technological obsolescence
    ( ) Insufficient policy or plan for preservation
    ( ) Insufficient resources for preservation
    ( ) Other—please specify:

17. What methods does your institution plan to use over the next 3 years to increase the level of staff expertise with digital preservation? (check all that apply)
( ) Local courses in computer or digital technology
( ) Training provided by professional organizations
( ) Training provided by vendors
( ) Independent study/assessment
( ) Hire staff with digital knowledge or experience
( ) Hire consultants
( ) Other—please specify:

18. Which of the following digital archiving services might your institution use if they were available at a reasonable cost (check all that apply)
( ) Technical training
( ) Policy considerations / recommendations (i.e., model policies)
( ) Administrative considerations (i.e., training in project mgt. and budgeting)
( ) Standards and best practices
( ) Consultant services
( ) Cooperative or shared storage/access/preservation facility
( ) Services provided by recommended 3rd party vendors, please indicate:
( ) Indexing
( ) Description
( ) Management of metadata
( ) Outsourcing of conversion projects
( ) Conversion or migration services
( ) Database development
( ) Maintenance
( ) Backup services
( ) Archiving or preservation storage services
( ) Access services
( ) Hardware / software to meet archival needs
( ) Other—please specify:

19. Name of Institution: _____
(Please indicate which repositories are included if your institution has digital holdings in more than on repository): _____
20. I am available to be contacted for further information to advance the purposes of this study.
( ) No   ( ) Yes
Name: _____
Title: _____
Phone: (   ) _____
E-mail: _____

21. **Comments:**

RLG Interview Questions

Name:
Title:
Institution:
Date/Time:

1. Responsibility for the preservation of digital materials is having an impact on many institutions. Would you please describe the changes or effects that this responsibility is…

   …causing for your institution?

   …changes/effects on you (professionally), your role?

2. Over the next 4 years, what do you consider to be the major (2–3?) issues facing your institution with regards to preservation of digital materials?

3. If it were possible for institutions to pool resources into a consortium or network, how would you see your institution fitting into such a network?

   What resources would you like to have available for your use?

   Who should manage or coordinate such a network?

4. Have you had any experience in selecting or using 3rd party vendors for outsourcing of digital preservation services?

5. What role does campus computing services play in helping to meet your needs for…

   …controlled storage?

   …other services?

6. With regards to digital preservation, if you could choose one concern or problem to be immediately and magically solved, what one would you choose?

7. Before we end, are there any other areas of digital preservation which you would like to bring to RLG's attention?

Have you completed the RLG survey on digital preservation?
( ) YES—If yes: thank you
( ) NO—If no: we'd like to request that you participate in the survey as well, to ensure complete representation in our analysis. RLG will forward a survey to you via e-mail within 24 hours.

Verify e-mail:

Thank you for taking the time to share your views.

Appendix B:
List of Participating Institutions

Amon Carter Museum
Boston University
Brigham Young University
British Library of Political & Economic Science,
  London School of Economics
Brooklyn Museum of Art
Canadian Centre for Architecture
Center for Research Libraries
Chicago Historical Society
Cleveland Museum*
Columbia University Libraries—Academic
  Computing*
Cornell University*
Duke University*
Florida State University
Folger Shakespeare Library
Harvard University
Hebrew Union College
Institute for Advanced Study
International Institute of Social History
Kimbell Art Museum
Jewish Theological Seminary
Maryland Historical Society*
Massachusetts Archives
Minnesota Historical Society*
Monterey Institute of International Studies
National Archives and Records Administration
National Library of Australia
National Library of Wales

Naval Air Warfare Center Technical Library
New York State Archives and Records
  Administration
New York State Library*
Pennsylvania State University
Princeton University
Public Record Office, UK
Rutgers University Library*
Smithsonian Institution
Smithsonian Institution Libraries
St. Louis Art Museum
State Historical Society of Wisconsin*
State University of New York - Albany
State University of New York - Stony Brook
Swiss National Library
Trinity College Library, Dublin, Ireland
University of British Columbia East Asian
  Library
University of Chicago Library*
University of Florida
University of Hawaii at Manoa
University of Leeds
University of Newcastle-upon-Tyne
University of Oklahoma
University of Oxford*
University of Rochester
University of Washington Law Library
Yale University Library*
Yeshiva University

* Staff from these institutions participated in interviews. In addition we interviewed staff from the Alabama
  Department of Archives and History; the University Library, University of Michigan; and the Bentley
  Historical Library, University of Michigan.