

Jason M. Casden. Project to Program: A Brief Study of Current Digital Preservation Practices in Small to Medium-sized Cultural Heritage Institutions. A Master's Paper for the M.S. in L.S. degree. April, 2006. 45 pages. Advisor: Paul Conway

This study sets out to contribute to the understanding of current practices in digitization projects at small and medium-sized cultural heritage institutions by interviewing digitization project managers at four of these institutions and analyzing their responses as case studies. The overarching theme of this study is the necessary transition from project to program at these institutions, and what these institutions are doing to make this transition. Additionally, this study looks at digital storage media and repository management at these institutions, which is an under-documented area of current practices. Through the use of a literature review and four case studies, several recommendations are made for improving the state of the art of digital preservation at these institutions and preparing them for the transition of digitization projects to digital programs.

Headings:

Electronic data archives -- Conservation and restoration

Electronic data archives -- Administration

Preservation of library materials -- Automation

Virtual library

PROJECT TO PROGRAM: A BRIEF STUDY OF CURRENT DIGITAL
PRESERVATION PRACTICES IN SMALL TO MEDIUM-SIZED CULTURAL
HERITAGE INSTITUTIONS

by
Jason M. Casden

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2006

Approved by

Paul Conway

TABLE OF CONTENTS

Introduction.....	2
Literature Review.	2
Research Question.	20
Methodology.....	21
Case Studies.....	22
Discussion.....	32
Conclusion.....	36
Bibliography	38
Appendix A. Phone Interview Protocol.....	41

Introduction.

This is a study of the transition from project to program for digitization projects at small and medium-sized cultural heritage institutions, and what these institutions are doing, if anything, to make this transition. In order to get a better understanding of these institutions' preparedness for this transition, this study presents four case studies showing current practices. Additionally, these case studies are supported by a literature review discussing the issues associated with this transition.

A particular interest of this study deals with storage media, including methods for coping with technological obsolescence, costs involved and organizational changes needed when considering preservation-quality digitization. Given the steep costs often associated with commercial preservation-quality digitization and the maintenance of such data, I feel that organizational changes are needed in order to reign in costs for smaller institutions, using in-house, outsourced or consortial repository solutions. In order for this to be successful, a better understanding of the nuts-and-bolts of repository storage management is needed.

Literature Review.**1. Project to program**

“As we begin the 21st century, however, libraries and archives face a critical transition in which digital projects must give way to digital programs to survive. In other words, institutions must come to terms with the digital collections they develop. Projects by their very nature are of limited duration and scope, most often involving efforts to create digital resources. Programs are ongoing and encompass the full life

cycle of digital resources, from selection and creation to management, access, and preservation.”

Anne R. Kenney, “Projects to Programs: Mainstreaming Digital Imaging Initiatives,” 2000.

In “Projects to Programs,” Anne Kenney emphasizes the need to transition from projects to programs in order to “safeguard” the “long-term value and utility” of digital collections as “institutional assets.” Under this premise, she outlines several elements of this transition, including: “devote institution resources to the transition,” “develop formal policies to encompass the life cycle of digital resources,” “tie policies to institutional resources,” “analyze current digital imaging projects to efficiencies and economies,” and “document the process and product,” among others. (Kenney, 2000) The overall themes of the elements deal with secure funding and institutional support for stable operations and increased field-wide knowledge of best practices and other digital preservation strategies. Kenney’s recommendations are practical, sound and echoed in other literature.

The 2001 Council on Library and Information Resources (CLIR) report *Building and Sustaining Digital Collections* also emphasizes the need for more cohesive, long-term, programmatic solutions to the issues associated with digital collections. “Given the challenges posed by the Web, and the new public that it brings to libraries and museums, the focus on addressing short-term problems with short-term solutions risks undermining the public trust these institutions have earned over time.” (CLIR, 2001) In discussing what is needed to sustain a digital project, the CLIR report focuses heavily on the need for a better “knowledge environment,” that includes “standards and best practices” and “ongoing means for creating and sharing a knowledge base” to serve as “a central clearinghouse for technical information.” In addition to an improved “knowledge

environment,” the CLIR report also emphasizes the need for “coherent and common digital architectures,” noting that “many conferees also advocated the development of digital service bureaus to provide an array of conversion, distribution and archiving services for a variety of cultural repositories.” They balance this suggestion with a question about the “feasibility” of such centers.

In their 2002 study *The Digital Library: A Biography*, Daniel Greenstein and Suzanne E. Thorin echo others’ calls for improved long-term digital preservation programs.

“Having acquired core competencies and technical understanding, the maturing digital library abandons the “build it and they will come” philosophy that characterized earlier approaches to collection development. It focuses instead on integrating digital materials into the library’s collections and on developing (and supporting with core funding) the policies, technical capacities, and professional skills needed to sustain it.” (Greenstein and Thorin, 2002)

They go on to identify several characteristics of “maturing digital libraries,” beginning with an “interest in modular architecture.” These libraries understand the complexity of the systems needed to support their collections. Additionally, the “desire for common standards” (Greenstein and Thorin, 2002) is echoed throughout the literature in calls for better best practices, shared knowledge and communication; something CLIR refers to as an improved “knowledge environment.” (CLIR, 2001) Additionally, they see more of a “focus on the user,” arguing that “the maturing digital library also seems to rediscover users.” In particular, they refer to the tendency of libraries in an “experimental phase” of digital collection development to turn a blind eye to user needs in favor of technological exploration.

The “interest in modular architecture,” relates to their argument that changes must be made to how a “maturing digital library ... is sustained technically and

organizationally.” Maturity relates both to pure technical support, which can include organizational budgeting, and to in-kind support for campus departments. The authors also argue for the maturing digital library’s “organizational integration into mainstream library services,” an issue that can be seen in the case studies presented later.

Organizational integration is important to moving from a project to a program, making it easier to argue for permanent organizational funding, as well as improving a project’s ability to make use of other organizational resources to improve or promote the project.

2. What should a digital preservation program look like?

In her 2001 CLIR publication *Strategies for Building Digital Collections*, Abby Smith emphasizes the long-term planning concerns of other authors. Smith argues that “Any assessment of what libraries have achieved so far must take into account two key factors common to sustainable collection development, be it of analog, digitized, or born-digital materials,” which are “a strategic view of the role of collections in the service of research and teaching” and “life cycle planning for the collections.” (Smith, 2001) The life-cycle planning theme is something that occurs across the digital collection and preservation literature, and is a good fit as a central theme for evaluating digital preservation programs. Smith goes as far as to say that “ensuring long-term access to digital collections depends on careful life-cycle management.”

Smith voices several concerns about the transition from project to program.

“How does the library budget for not only the creation of the digital scans but also for the metadata, storage capacity, preservation tools (e.g., refreshing, migration), and user support—the sorts of things that are routinely budgeted for book acquisitions? How much of the program is supported by grant funding and how much by base funding? If the program is currently grant supported, what plans exist to make it self-sustaining?” (Smith, 2001)

At the same time, she does leave room for projects which may not demand extensive long-term support, instead emphasizing an assessment of project needs, rather than total long-term support at all times, saying, “A sustainable digitization strategy may well include the creation of digital surrogates that serve short-term needs and do not demand long-term support. The crucial thing is to anticipate what support, if any, will be needed.” (Smith, 2001)

William LeFurgy, in his 2002 article “Levels of Service for Digital Repositories,” identifies three “levels” of digital collections: “optimal, enhanced, and minimal.” (LeFurgy, 2002) In his paradigm, digital collections are rated by their ability to hold “fully persistent” digital materials and “make them available to users.” While the extreme ends of this scale may be difficult to design, the case studies I will examine later span a reasonably wide range of preparedness for the persistence of digits as it relates to storage media.

In describing his “levels of service,” Lefurgy says that the different levels of service for a digital collection “can best be thought of as a matrix with one set of values determined by the available technology and the other set determined by the degree to which digital materials have persistent qualities.” (LeFurgy, 2002) In this model, persistence independent of technology may be achieved by conforming to “exacting rules,” aided by concepts like the Open Archival Information System (OAIS) reference model, although LeFurgy notes that materials that conform to such rules “will constitute only a small fraction of the overall universe of information.” In regards to technological support, LeFurgy sees three “Phases” of technological tools and processes, with Phase I using “rudimentary tools and processes ... obtainable today” (at the time of this writing,

2002), Phase II using “more advanced tools and processes ... obtainable over the next decade” and Phase III “very advanced tools and processes ... obtainable after the next decade.” Although most commentators are now more concerned with what can be done with currently obtainable technology, LeFurgy’s Phases can be seen across a range of currently available technologies. For example, Phase I may represent the storage of digital materials solely on CD-ROMs, while Phase II may represent the storage of digital materials on CD-ROMs and PC hard disk drives and Phase III may represent the use of server-quality hard disk drives with RAID redundancy schemes and offsite tape backups. Regardless, LeFurgy notes that “getting large quantities of materials into the enhanced and optimal service levels will require dramatic change in how digital materials are now produced and maintained.”

In their 2003 paper “Comparing Library Resource Allocations for the Paper and the Digital Library,” Lynn Connaway and Stephen Lawrence conduct a study of eleven librarians at Association of Research Libraries (ARL) member libraries, gathering their impressions of the differing cost environments of “all-paper” and “all-digital” libraries. (Connaway & Lawrence, 2003) This study provides an interesting view into the current state of knowledge of librarians as it relates to digital storage, and will do well to set the tone for this paper.

Connaway and Lawrence begin with a solid argument, saying that “rapidly evolving information technologies and media options provide substantial motivation for librarians to understand library cost structures so as to make informed decisions about acquiring and utilizing new technologies.” (Connaway & Lawrence, 2003) This is in line with my own view, that the better understanding they refer to is necessary for all digital

storage management scenarios. When outsourcing to OCLC or another company, a better understanding of the cost structures and technological possibilities will improve a librarian's ability to shop for a repository that provides the most reliable preservation with the most efficient cost structures. For institutions planning in-house storage management, or creating consortial repositories, this knowledge is even more crucial. Fully understanding the financial and technological aspects of storage management is vital to maintaining a repository that neither fails as a preservation technique, nor expends resources too wastefully.

Unfortunately, the results and conclusions of this study may not agree with more concrete cost studies, such as Stephen Chapman's study. (Chapman, 2003) Connaway and Lawrence enter the study with four explicitly stated assumptions: fewer "labor," "space," and "material" resources will be needed in an "all-digital" library as compared to an "all-paper" library. They also assume that "equipment resources," which includes digital storage media, will be approximately the same in each type of library. The perceptions of the eleven librarians mostly support these assumptions, with the study indicating "an agreement among the participating librarians that labor, aggregate space requirements and material resources are estimated to be less in an all-digital library than in a paper-library." Furthermore, the librarians in the study go as far as to estimate that even commitments of "equipment resources" will be fewer, an assumption even the authors were leery of. (Connaway & Lawrence, 2003)

3. Challenges in creating an archival master

"Given these choices, I would submit that microfilm, which is durable as a means of preserving content but hard to use, is not the obvious choice as a preservation

technology when compared to digital imagery, which must be regularly renewed but which promises to be relatively easy to use and therefore an effective means of preserving access.”

Donald J. Waters, *Electronic Technologies and Preservation*, 1992.

In his paper “Electronic Technologies and Preservation,” Donald Waters takes an optimistic view of digital preservation. Early in his work, he proposes six “enabling principles” of digital preservation, which are “think in terms of life cycles, not permanency,” “simplify,” “adopt an incremental approach,” “formulate working (and testable) hypotheses,” “build technical activities on standards and products being developed for the broad marketplace,” and “cooperate to make digital image documents widely accessible.” (Waters, 1992) Underlying all of these enabling principles is an assumption that a reasonable attempt at digital preservation is possible with technology that is available now. The principle of the life cycle mindset may be the most useful when discussing digital storage media. Rather than trying to hold back the flood of digital information until the lifespan of storage media multiplies ten-fold, it may be more useful to accept that permanency is not an appropriate paradigm when considering storage media. As Waters writes, “we must now think deliberately in terms of a relatively short renewal period.” (Waters, 1992) Rather than proposing future innovations to an industry over which libraries have little influence, Waters says that we need to develop “good cost models” for working with digitized documents. This work, combined with his later report for the Research Libraries Group, seems to indicate an ongoing need for incremental improvement in digital preservation program support that may not reach a stable endpoint for some time.

Paul Conway's popularly cited paper "Preservation in the Digital World" presents a balanced look at the issue, with some very interesting comments regarding digital storage media. He provides us with a fascinating graph, showing how almost invariably as character density increases, storage media lifespan decreases. The graph begins with clay tablets, and ends with optical storage, and features a lifespan blip (cited here as 300 years) at the development of microfilm. (Conway, 1996) This graph makes it startlingly clear the path that media longevity seems to be on.

Conway emphasizes the need for organizational change to accommodate new technology, which seems in line with Waters' call for better cost models. He says that libraries "need to recognize their role in the development of digital imaging technologies, as well as the new demands that these technologies will place on them as organizations." (Conway, 1996) He also makes reference to the idea that the struggles libraries are having over digital preservation may mirror similar and earlier struggles in the corporate world.

Of the many useful concepts set forward by Conway in this paper, the most directly applicable to digital storage media are his assertions to both "protect original items" and "maintain digital objects." While this may be excellent advice, in a situation with tight organizational budgets, protecting original items may fall by the wayside due to the (perhaps false) comfort of digital copies. Conway discusses the idea that a chain of technology only being as strong as its "weakest component," saying digital storage media "most likely will far outlast the capability of systems to retrieve and interpret the data stored on them." Finally, Conway emphasizes the importance of hardware migration,

especially considering his understandable nervousness on the issue of media longevity. (Conway, 1996)

These views are similar to the recommendations of Michael Lesk in his report “Preservation of New Technology,” where he says that “in this new world, preservation means copying, not physical preservation.” (Lesk, 1992) Despite echoing, in much more detail, much of the anxiety the other authors held regarding conversion of file formats and storage media, he tries to counter the weight of this anxiety by saying that “two hundred years after the invention of the metric system it is still not universally accepted, we must accept that conversion will always be with us.” By stating this, it seems to give credence to the idea that we need to move forward with our best options, despite the hopes and fantasies of better solutions over the horizon.

Lesk also approaches the issue of the organizational changes needed to accommodate successful digital preservation. He says that since digital preservation means copying, “it requires more attention to long-term costs.” (Lesk, 1992) Organization planning and budgeting will need to be changed significantly for the endless migrations necessary for digital preservation. Nevertheless, in his 1996 paper “Substituting Images for Books: The Economics for Libraries,” Lesk, using numbers from Waters, provides aggressive predictions that costs of “digital and traditional library operations” will “cross over in about 5 years,” and “in 10 years electronic storage has a major cost advantage.” (Lesk, 1996) Earlier we saw Chapman (2003), seven years after this prediction, argue that the current numbers seem to contradict Lesk’s extrapolated numbers, in at least some cases.

4. Current practice

In their very well-known 1996 paper for the Research Libraries Group “Preserving Digital Information: Report of the Task Force on Archiving Digital Information,” Donald Waters and John Garrett seek to make recommendations for improvements in digital preservation practices based on current observations, with a particular focus on “data migration.” (Waters & Garrett, 1996) Waters and Garrett expand the notion of what it means to accomplish digital preservation beyond the emphasis on creating archival masters found in Waters’ earlier work (1992). Based on their observations in 1996, they argued that “the first line of defense against loss of valuable digital information rests with the creators, providers and owners of digital information,” reinforcing arguments for improved life-cycle management and communication between different data custodians. Additionally, they suggest the possibility of “certified digital archives,” a possible need that was also identified by the earlier CLIR study. (2001) Finally, they found that “long-term preservation of digital information on a scale adequate for the demands of future research and scholarship will require a deep infrastructure capable of supporting a distributed system of digital archives.” While the full realization of this argument may take awhile, the direction of technological improvement in digital preservation infrastructures is a good one to follow. (Waters & Garrett, 1996)

Waters and Garrett go on to argue for the improved “knowledge environment” that was echoed in CLIR (2001), saying that “the preparation of white paper on the legal and institutional foundations for the development of effective fail-safe mechanisms” is needed, and that “follow-on case studies” are needed to “identify current best practices

and to benchmark costs,” a finding that I hope to support with my case studies. (Waters & Garrett, 1996)

In the 2004 study by the PREMIS Working Group, researchers conducted case studies to determine the “current practice” for “a variety of areas, such as mission, funding, preservation strategy, and access policies.” (OCLC & RLG, 2004) While the study primarily focused on “managing preservation metadata,” it drew some general conclusions from its case studies that are also relevant in this paper. The broadest and most troubling claim is that “the cultural heritage community has very little experience with digital preservation.”

Among the claims most relevant to this paper is the claim that while “90% of respondents funded their repositories from their operational budget ... two-thirds used internal or external grant funds, in addition to, or instead of, operations funds.” (OCLC & RLG, 2004) This ratio is even more severe for the four case studies included in later sections. A more optimistic finding is that “the majority of institutions chose more than one strategy for preservation,” a decision that the OCLC & RLG study encourages. Furthermore, the researchers recommend maintaining multiple versions of files in the repository, something which came in very handy in Case Study #1 below.

Stephen Chapman, in his study “Counting the Costs of Digital Preservation: Is Repository Storage Affordable?” begins with several bold statements about digital repositories and preservation deduced from his studies of the actual costs of repository storage. He argues that “the repository is the nucleus of preservation activity,” saying that “OAIS implies that without a repository, preservation is not possible.” Secondly, Chapman argues that “most cultural heritage institutions are likely to be consumers of

centralized preservation services, rather than architects and managers of digital repositories.” Underlying the entire study is his thesis that “repository storage costs and payment schedules (independent of costs for ingest or access) must be affordable and manageable.” (Chapman, 2003)

This article compares physical media storage costs at the Harvard Depository against digital information storage costs at OCLC’s Digital Archive. Chapman looks at four different data formats, namely “Text,” “Photos,” “Audio” and “Moving Images,” with several subcategories for different images qualities or storage media. (Chapman, 2003) This study provides an interesting contrast to Connaway and Lawrence’s study of librarian perceptions of digital repository costs. Chapman found that, contrary to the perceptions of librarians (Connaway & Lawrence, 2003), digital storage was significantly more expensive than the physical repository storage in most cases, and often dramatically so. The one case in which digital storage may be cheaper than the physical repository was for ASCII text when compared to microfilm or books. Unfortunately, ASCII is a fairly poor representation of a book. In contrast, “the most favorable digital-to-analog cost gaps ... are 1.52:1 for page images versus books, and 2.47:1 for page images versus microfilm stored in the HD film vault.” (Chapman, 2003) Chapman uses this data to criticize Lesk’s (1996) assertion that “the costs of the digital and traditional library operations [would] cross over in about five years.” (Chapman, 2003)

5. Planning

In their 2000 report on a study for the Joint Information Systems Committee (JISC), “A strategic framework for creating and preserving digital resources,” Daniel

Greenstein and Neil Beagrie use case studies to identify some elements of a framework that may be used to achieve the kind of long-term persistence that concerned LeFurgy. They begin by saying that “the great advantages of digital information are coupled with the enormous fragility of this medium over time compared to traditional media such as paper,” describing electronic information as “fragile and evanescent.” (Greenstein & Beagrie, 2000)

Like other authors, they emphasize the need for “best practices,” but also emphasize the difficulty of enforcing uniform best practices at every stage in the life cycle of digital materials, because “different (and often, differently interested) stakeholders become involved with data resources at different stages.” As a result, “decisions which affect the prospects for and the costs involved in data preservation are distributed across different ... stakeholders.” Additionally, they found that “there were few channels established to facilitate ... inter-communication” among the various stakeholders. For very small archives, communication is an even greater concern, as data storage and other IT tasks are often outsourced either to the University IT department or to external contractors. If archives and libraries do not take measures to ensure “inter-communication” with other stakeholders about best practices for data preservation, data loss or corruption is a significant possibility.

An important element of digital preservation planning that is being focused on here is digital storage media migration. Conway (1996), Lesk (1992) and Waters (1992) were shown above to be concerned with this issue, and have emphasized migration as an important component of a digital preservation program.

6. Risk of inaction

“We would like to have a reliable storage life of the machine-readable data at least as long as archival microfilm, which is presently about 50 years. This is based on the cost analysis, which shows inspection, manual handling, and recopying as being the major cost elements in data storage.”

Gerald J. Rosenkrantz, “National Archives Mass Storage Requirements – 1975 to 1980,” December 1971.

This sentiment is echoed by many of the digital preservation theorists writing today, a third of a century later. Despite this, digital storage media today are not significantly closer to the kinds of lifespan that Rosenkrantz and other archival writers desire. Paul Shields says that magnetic tape has a life span of 1-3 years, with hard drives at 3-5 years and optical disks at 30 or more years. (Shields) This also does not consider reliability across individual digital storage media and, as Eaton says, quoting Ron Kushnier, “all CD-ROMs are not created equal.” (Eaton, 1994) Although the life span issues of digital storage media have not improved significantly in the last 34 years, the use of and dependence on these media by libraries and archives has increased at a blistering pace nonetheless.

In “Electronic Media and Preservation,” Fynnette Eaton of NARA takes a very practical approach to the issue of digital storage media preservation and management. She argues that seven elements of a traditional preservation program, “environment, storage, handling and use, microreproduction and reformatting, exhibition, disaster planning and treatment” all must also be considered with electronic records. (Eaton, 1994) Connecting traditional preservation with the “new” science of digital preservation is very useful, and may help to bring the issue down from the heights of hysteria back down into the realm of manageable challenges. Eaton also emphasizes environmental

controls, calling it the “single most important factor in the preservation of electronic media.” (Eaton, 1994)

The author seems very unconcerned with the issues associated with reformatting, saying that the “recopying process simply reformats the information to avoid obsolescence.” (Eaton, 1994) Eaton has some optimism that optical storage media, like CD-ROMs, may bring a longer lifetime to the table for digital storage media, but laments that lack of standards or dependability among individual CDs. Again emphasizing to start taking measures towards better digital preservation sooner rather than later, Eaton invokes the familiar preservation concept that “the best method of treatment is prevention,” saying that archives should recopy electronic files “before serious problems develop.” (Eaton, 1994)

In her work “Digital Preservation: A Time Bomb for Digital Libraries,” Margaret Hedstrom paints a fairly intimidating picture of digital preservation, arguing that “the two terms “mass storage” and “long-term preservation” embody a contradiction in the current state of affairs of digital library development, representing a time bomb that threatens the long-term viability of this new type of knowledge resource.” (Hedstrom, 1998) She states that “digital preservation is constrained by the absence of established standards, protocols and proven methods for preserving digital information and by the tendency to consider preservation issues only at the end of a project or after a sensational loss.” (Hedstrom, 1998) Of particular interest is her discussion of the issues associated with digital storage media. She sets out ideals for storage media, which include “a long life expectancy, a high degree of disaster resistance, sufficient durability to withstand regular use, and very large storage capacities,” as well as “modest” equipment, media and maintenance costs.

(Hedstrom, 1998) In her discussion of problems with storage media life spans, she twice cites microfilm as having a potential lifespan of “at least 500 years with minimal maintenance if stores properly,” which is a considerably higher number than the 50 years cited by Rosenkrantz in 1971. (Hedstrom, 1998) After setting a bar for storage media that is distant in, depending on your level of optimism, the future or ones imagination, Hedstrom moves on to criticize some of the methods popularly considered to offset issues with digital storage media lifespan.

Hedstrom argues that migration from obsolete systems is more difficult than it seems, and is not simply an issue of “transferring a stream of bits from old to new media” because “complex and expensive transformations of digital objects often are necessary.” (Hedstrom, 1998) Hedstrom seems to be combining file format migration and hardware migration into a single and inextricably linked issue. This seems to only muddy the waters, limiting the ability to solve the problem that “there is limited experience with the types of migrations needed to maintain access to complex digital objects over extended periods of time.” (Hedstrom, 1998) An additional issue with that lack of experience is one of looking outside of the field; these sorts of issues have been considered, with much larger budgets, in government and corporate settings for decades. While file format migration is a very difficult issue, especially for born digital objects, hardware migration may at least keep these bits alive until a file format migration method can be devised. Furthermore, with digitized objects, standard and/or popular formats, like TIFF, are often used, which may simplify the file format issue in the future (at least for non-born-digital objects).

Hedstrom cites Conway's statement that digital preservation depends on the weakest technological link. Given the amount of data that now lives primarily in a digital world, I would think that this would prompt a call for a "best we've got" solution to preservation, instead of the technological brainstorming that happens in this paper. Hedstrom mentions distant solutions that range from wishful thinking to science fiction, such as "the development of backward compatibility paths that would be included as a standard feature of all software," and the future possibility of using "LANL Ion Beam Storage" to create storage media with a longer lifespan. Unfortunately, the problem exists now, and must be treated with what we have available to us, even if all we have are leeches and tourniquets. (Hedstrom, 1998)

In his paper "Ensuring the Longevity of Digital Information," a revision of his popular article "Ensuring the Longevity of Digital Documents" from *Scientific American*, Jeff Rothenberg presents another frightening view of the future of digital information. The narrative is set in the year 2045, with Rothenberg's grandchildren trying to recover data critical to their grandfather's estate off of a CD-ROM from 1995. Rothenberg uses complicated in-depth explanations of bit structures, encodings and interpretation to plunge the reader into an unfamiliar and uncertain world of complexity on the edge of human comprehension. He infuses his story with emotion by evoking the lines of Shakespeare's 18th sonnet. (Rothenberg, 1995)

Rothenberg's work reflects many of the same anxieties found in Hedstrom's piece. He is concerned with the short lifetime of digital media, which he estimates at about five years after taking into account the weakest technological link concerns. Additionally, he is also critical of the idea of data migration (as is evident from his

section heading “the false promise of migration”), writing that “like an illiterate monk dutifully copying text in a lost language, migration may save the bits but lose their meaning.” (Rothenberg, 1995) Like Hedstrom, he conflates file format migration with physical data migration and, in doing so, seems to make the situation seem so daunting as not to be worth attempting at all.

Unlike Hedstrom, he is also very critical of the idea of standards being able to ameliorate the problems associated with digital preservation. In the sections of his paper “The illusion that standards provide an answer,” Waters uses the complicated example of Relational Database Systems. His argument is that despite standards that specify how a Relational Database System should function, different implementations function in differing and often non-standard ways. This example is problematic because it rejects the idea of a technological critical mass. While standards are useful for homogenizing like kinds of data, at least to a certain extent, when choosing technologies for special collections, it is also important to choose a technology with a critical mass of archival use (even if it is not based on a standard).

Research Question.

Using the recommendation of additional “follow-on case studies” of current practices in digital preservation by Waters and Garrett (1996) as a starting point, this study will examine how prepared smaller cultural heritage institutions are for transitioning from digital projects to programs. Additionally, how are small to medium-sized institutions coping with the relatively new demands associated with digital repository management and digital preservation when conducting digitization projects? I

expect to find a lack of preparedness similar to that found in the PREMIS study (OCLC & RLG, 2004), as well as a wide-range of solutions based on available funding and staff. Finally, I am interested to estimate the impact of the abundance of “best practice” documents in some areas of digital preservation, as well as the lack thereof in others.

Methodology.

1. Goals and presentation

This study is intended to broaden the understanding of current digital preservation practices, particularly at smaller archives which are conducting digitization projects. The study was conducted with the following general themes as a centerpiece:

- What are the current digital preservation practices at these institutions?
- What are their current long-term digital preservation plans?
- What kind of issues have they had related to data preservation?
- What kind of support system, financial and otherwise, is needed for digital preservation programs at these institutions?

Additionally, there is an overarching concern for storage media issues and physical bit preservation.

This study’s data consists of responses to four interviews, presented as individual case studies. Each study provides an interesting perspective on current practices, and as a whole they display a wide range of digital preservation preparedness and planning.

2. Subjects

Subjects were recruited through a posting on the “Archives & Archivists” electronic listserv. Six responses were received, and four of these respondents were able

to schedule interviews. Each respondent was emailed a general confidentiality disclaimer along with an “issue outline” before scheduling a phone interview. Respondents were asked if they were available for follow-up questions, and all agreed. The level of enthusiasm and helpfulness from the subjects was very high.

3. Information gathering

Phone interviews were conducted using a loose questionnaire of about a dozen questions, depending on how earlier questions were answered. The interview protocol is included as Appendix A. Respondents were so thorough and eager in their answers that they often answered several unasked questions with a single answer. Respondents were not provided with the questions ahead of time, but the questions did not stray thematically from the “issue outline” that was emailed to them. The interviews lasted approximately thirty minutes each. The case studies were then written from interview notes with as similar a thematic structure as the data would allow.

Case Studies.

Case study #1: A state-wide cultural heritage institution.

1. Institutional Profile

This project is a state-wide cultural heritage digitization project, which functions under a northeastern state-wide, state-funded historical society. This historical society is home to a library and a museum, and was founded in the first half of the 19th Century. With over 25 staff, they provide these services to residents of the state, in addition to educational programs.

2. Project Profile

The digitization project has been funded completely by private and public grants since it began in 1999. The original materials include photographs, glass-plate negatives, hand-written manuscripts, letters, postcards, daguerreotype, paintings and maps. No audio or video content is being digitized, but some audio narrations and 3-dimensional object viewing tools have been created for the project. Additionally, the project staff instructs smaller state cultural heritage institutions on how to digitize their own materials. These images are housed with the rest of the images in this project, although the individual organizations are responsible for maintaining the master images.

3. Institutional Support

The project has received strong support from the larger institution, which views the project as the public face of the organization and has included it as part of the large institutions strategic plan. The project is viewed both as an outreach tool to the public and to the other cultural heritage organizations scattered throughout the state. As grant funding comes to a close, it is felt that the institution will do its best to pick up the slack.

4. Digitization

The digitization was all done in-house using flat-bed scanners and digital cameras for image creation, and Adobe Photoshop for image processing. Images are created in TIFF and JPEG formats; the 300 dpi, 24-bit TIFF images are considered the master copies, while the low (10%) compression JPEG images are for web viewing. The recommendations produced by the “Collaborative Digital Program” at the University of Denver were consulted when determining these specifications. Once the JPEG images are uploaded to the web site, ImageMagick tools are used to create multiple viewing versions. The largest JPEG images are stored on a separate “media server,” while the

smaller variants generated by ImageMagick are stored on the web server. The smaller images are free to view, while the largest JPEG is can only be viewed after paying a fee.

5. Repository

The public web repository hosting is outsourced, which costs approximately \$6000 per month, funded by grant awards. Additionally, a service contract with the same company for the in-house systems costs about \$3000 per year. Finally, \$36,000 per year is spent on external web designers and programmers. The total yearly cost for these services is about \$45,000, and does not include hardware costs. Service gaps for the public web interface have been minimal, but there have been ongoing issues with download speeds. No systems staff was hired directly by the organization for this project, and the bulk of systems work is outsourced.

6. Backups

Backups are done in-house by both storing the images on internal servers with a RAID subsystem, and by copying all of the images nightly to external hard disk drives. Until recently, these backups were only performed for the images produced in-house, not for those produced by the smaller cultural heritage institutions that are feeding images into the wider project. These organizations were responsible for their own backups, and their images represent about a third of the images available to the public through this project.

7. Other Issues

In 2004, however, there was a total hard disk failure on the media server. This led to the loss of all of the “large” original JPEG images, about 9000 in all. Of those, only about 6000 were being backed up internally. As a result, the project staff had to contact

the smaller institutions in an attempt to restore the other 3000 lost images. Some of these images had not been backed up by the smaller organizations either, and were permanently lost and had to be rescanned. The entire recovery time was about six months. The smaller freely-accessible images created by ImageMagick were not lost, since they are stored on the web server. As a result, the disturbance was not noticed by most of the public. The project staff responded by beginning off-site backups to external hard disk drives of all images, protecting them more effectively from a future crash.

Case study #2: A religious and institutional archives.

1. Institutional Profile

This project is taking place at a religious and institutional archives in the Midwest. The archives has about a dozen staff members, and functions under the umbrella of a religious college. The archives was founded in the mid 20th Century, and focuses on preserving and communicating religious, cultural and institutional history.

2. Project Profile

The materials being digitized make up the records for a global religious institution located in the United States. The selected collection is composed mostly of correspondence, meeting notes and other text-based institutional records, with a small amount of audio and photographic materials. The collection is from the 20th century, bulk 1940s and contains 488.4 linear feet of material. The project began in 1999, and about half of the materials have been digitized. The digitization project is funded largely from both private and public grants, while the larger organization provides about 25% of the funding.

3. Institutional Support

The Project Director feels that the institution is very supportive of this project and its funding needs. The larger institution has the resources to support the project as it extends beyond grant funding and attempts to become a program.

4. Digitization

The digitization for this project has been entirely outsourced to a local company. The company picks up 20 boxes per month and produces digital images and microfilm for each item. The company produces 200 dpi bitonal Multi-Page TIFF images. While the Project Director remembers some ISO standards being looked at during the onset of the project, the origin of these scanning standards is unknown. The project staff have been happy with the company's work, both in regards to the end-product and the low damage to the original materials.

5. Repository

The bulk of the images are not available on the web. A few sample images are made available, but the project is not currently concerned with making the web-accessible. Copyright concerns were cited in the interview, but a lack of a systems staff and other resources are also a limiting factor. There is a full-time systems administrator in the department, but supporting this project is only one of many of this person's responsibilities.

6. Backups

The digital images are stored on CD-ROMs. There are three copies of each CD-ROM: one master copy, one user copy and one copy that is mailed to a national museum. Microfilm is used as a backup to the CD-ROMs. The CD-ROMs are checked once a year,

costing about 40 hours a year, but there is currently no schedule for an overall media migration. The project director said that they are currently waiting to see what the next technological development will be. Since the project began receiving CD-ROMs in 2000, none have yet needed to be replaced.

Case study #3: A university popular music archives.

1. Institutional Profile

This project takes place at a popular music archives at a large southern State University. The University, which was founded in the early 20th Century, has the largest student body of any University in its state. The archives has five full-time staff members, and is able to draw on the technical staff employed by the University.

2. Project Profile

The collection selected for digitization is composed of about 4000 song broadsides from the last three centuries, bulk 1880s. These broadsides are primarily text with some engravings on six by nine inch sheets. They were not created with longevity in mind, and many are very fragile. The project had been funded by a federal grant from 2003 to 2005, with the institution providing matching funds with in-kind services. The project is, aside from some “tweaking,” completed, and the Coordinator is “very proud of the outcome.”

3. Institutional Support

While the Director of the archives is seen as supportive by the Coordinator, the University is “definitely not.” While the University is glad the project exists, they have resisted making it a line-item in their budget. Luckily, a portion of the budget comes from

a statewide organization outside of the University which has provided some funding support for the project. Otherwise, funding is allocated out of the pre-existing allotment for the general archives.

4. Digitization

Digitization for this project took place in-house, using flat bed scanners and Adobe Photoshop. Each broadside is scanned in as a 24-bit, 600dpi, 6"x9" TIFF image, and color 300dpi JPEG viewing versions are subsequently created from the TIFF master. The Project Coordinator attended the NEDCC "School for Scanning" in 2001, from which she learned about "best practices" for digital image creation in archives. Additionally, various digitization resources were consulted, produced by University libraries, the National Initiative for a Networked Cultural Heritage (NINCH), the US National Archives and Records Administration (NARA), the Library of Congress, and the Digital Library Federation.

At this point, the images are stored on a server dedicated to hosting image files and are served to the public through an internally-developed web interface. TIFF files are provided along with the JPEG viewing copies because this had originally been specified in the grant proposal; given another chance, the coordinator would have left this requirement out of the grant proposal. The additional overhead of serving such large files is unwanted.

5. Repository

The web site is hosted in-house, although the project does not have any specific systems staff. The University systems staff provide some support for this system as well as the image repository and cataloging database servers. This includes emergency help as

well as routine maintenance updates, totaling about 8 hours per month. It is estimated that the project staff spend about 16 hours per month on systems-related work. Despite the lack of full-time systems staff, there have been no serious service gaps for the web site. All of the servers make use of some form of RAID data redundancy. The servers were purchased in 2003, and there is no current plan for migrating to newer servers.

6. Backups

Tape backups are performed every weekday by archives staff. The entire collection fits on one tape, and the last 5 tapes are stored in-house until Friday. On Friday, the latest tape is exchanged with a tape from another library to provide for off-site backup storage. Additionally, all of the master TIFFs have been stored on Mitsui Gold CD-ROMs, on-site. The CD-ROMs were created as the items were scanned, and there is no media migration plan in place. Every year 10% of the CD-ROMs are checked for corruption; should any be found to be corrupt, all of the CD-ROMs will be replaced. This has not yet happened. The overall backup scheme, along with the RAID storage, has thus far protected the project from any data loss.

7. Other Issues

In regards to systems staff, the Coordinator says that not having a full-time Systems Administrator is “a real hole in our staff structure.” The archives is currently trying to fund a new full-time position that will spend 80% of their time on systems administration tasks, and 20% of their time on cataloging.

Case Study #4: An audio collection at a small private college.

1. Institutional Profile

This project functions under a small private four-year college in the Western United States, which was founded in the 1970s. The religiously-inspired college has an enrollment of between 1000 and 1500 students, and has been host to some very famous names in its humanities programs.

2. Project Profile

This project employs four FTE staff members. This collection is composed of audio tapes, mostly ¼ inch reel-to-reel, of poetry readings, lectures, workshops and other poetry-related events featuring often very notable poets at this small private college. The project began in 2002, and is funded by private and public grants. In all, grant funds represent about 90% of the projects funds, with 10% coming from the college.

3. Institutional Support

This project is being used as a seed to develop a larger archives program at the college. The project is autonomous and is not part of the library. The college is supportive of the project, but has not yet provided the kind of funding necessary to sustain an archives program after the grant funding runs out. The project has received national exposure, and the college appreciates the credibility it has brought to the academic reputation of the institution.

4. Digitization

The digitization was done by in-house staff at an on-campus studio and a leased audio mastering studio with identical equipment. The master audio files are 24 bit 44.1k WAV files. Audio tweaking was done using ProTools software, and 16 bit WAV files were stored on CD-ROMs as listening copies. Professional sound engineers were consulted to determine the master audio file specifications. Additionally, “best practices”

being developed by the “Collaborative Digitization Program” at the University of Denver were consulted. A typical file is 1 to 1.5 hours long, and is not tracked. Instructors at this college and others can ask for specific clips, which will be created on-demand.

5. Repository

Web hosting is provided at no cost to the institution by Brewster Kahle’s Internet Archive. As the audio files are created, they are uploaded to the Internet Archive and served to the public free of charge. The project manager has been very happy with the Internet Archive’s service, but is hoping to create some additional interfaces to improve search results and to provide more context for the materials. Service gaps have been minimal.

6. Backups

Originally, all of the files were stored on “Gold” CD-ROMs, but that practice has since been discontinued in favor of redundant RAID hard drives and magnetic tape. It was decided that hard drive storage was not much more expensive or labor-intensive, and is much more reliable and less risky. Audio files are stored on an Apple XServe 2.5 TB Raid5 system, which cost about \$10,000. There is also a second backup hard drive that cost about \$3000. Additionally, weekly tape backups are created and sent off-site. The project goes through about \$1000 of tapes per year. Additionally, about \$14,000 of staff time goes into the management of this storage.

7. Other Issues

This project employs a full-time Technical Manager, at a salary of about \$40,000. Of this position, the project manager says, “this is absolutely essential, I don’t think you could do a project like this without a Technical Manager.” The Technical Manager is a

jack-of-all-trades, managing backups, server hardware and software, storage management, PHP programming and web design. Keeping this position once grant funds expire is a major concern for the project manager, who is trying to get the salary picked up by the college.

Discussion.

1. Institutional Support and Funding

These case studies all displayed a very heavy reliance on external grant funds. This was to be expected, based on the OCLC (2004) study, which says that of the institutions they studied, “two-thirds used internal or external grant funds.” The institutions I looked at showed an even heavier dependency on grant funds, with all of them relying heavily of these funds for their operational budget. Additionally, none of the organizations studied had concrete plans for shifting their funding to internal sources. This is obviously a serious problem when trying to transition from a project to a program, as all of these programs are facing the imminent end of the bulk of their external grant funding.

The projects studied here diverge on the issue of perceived institutional support. There are two examples of strong institutional support, with the institution behind the project in Case Study #4 going so far as to make the project part of their strategic plan. There is one example of very weak institutional support, along with one example “lip service” support, with no concrete resources having been allocated or discussed. Case Study #4 is clearly in the best situation here, and is a model project in every category that is examined here. By becoming an explicit part of an institution’s strategic plan, the argument for current and ongoing funding becomes much more convincing.

2. Digitization

The digitization specifications are a significant area of general agreement among all of the case studies. All of the scanning projects used TIFFs as their master files, and the two working with color materials (Case Studies #1 and #3) scanned them with at least 300 dpi and 24-bit color, while the one working with institutional records (Case Study #2) scanned with 200 dpi. All of these projects then created high quality JPEG viewing copies. The audio project also used high quality specifications for the master files. Additionally, all of these projects, with the possible exception of Case Study #2, consulted “best practices” documentation, while some made use of additional resources, such as workshops and regional experts.

That four projects which differ so dramatically in repository management can be so similar when creating archival masters is telling. The availability of extensive “best practices” research and documentation, conference talks and regional scanning workshops have clearly had an effect on these smaller institutions. In contrast, the lack of such an extensive “knowledge environment” (CLIR, 2001) for repository management issues shows in the disorganized and often highly risky repository strategies of these organizations.

3. Repositories

The file repositories used by the projects in this study show a range of ongoing practices. Two of the projects outsource their web repositories, one hosts and manages their in-house, while one (Case Study #2) does not make use of a comprehensive file repository. Case Studies #1 and #4 manage an internal file repository on external disk drives with a RAID storage scheme for redundancy in addition to their external web

repository. Again, Case Study #4 shows itself to be very strong in regards to digital preservation, fitting well into the OCLC (2004) recommendation to “choose multiple strategies for digital preservation.”

The tendency to outsource repository services is noteworthy. Particularly with smaller projects, it can be difficult to fund a permanent Systems Administrator to manager repositories and backups in-house. At the same time, the Project Manager’s awareness of the technical standards being implemented in the off-site repositories seemed to be low. This argues for the development of a digital preservation certification for repository services, as suggested by Waters and Garrett (1996) and CLIR (2001).

In regards to storage media, three out of the four projects studied were using RAID for data redundancy for at least some part of their data storage scheme. Increased awareness of fairly easily and cheaply implemented systems like RAID storage will likely result in increased bit-level digital preservation at institutions with insufficient Systems staff.

4. Backups

Backups are another split issue among these projects, with only two projects (Case Studies #3 and #4) backing up data to tape. The use of hard disk drives for bit preservation was an alternative to tapes for Case Study #1, and a supplement to tape backups for the rigorous Case Study #4. The use of CD-ROMs as a backup medium seems popular, but is questionable from a preservation standpoint. Case Study #4 began by using CD-ROMs, but found a redundant hard disk drive system with tape backups not to be unreasonably more expensive relative to the increase in preservation functionality. Case Study #3 uses CD-ROMs to supplement their tape backup system, while Case Study

#2 relies on CD-ROMs entirely, without any scheduled media migration. This project is definitely standing on weak legs from a preservation perspective, and may one day find itself with some data loss, despite the off-site storage of multiple CD-ROM copies. CD-ROM life spans are difficult to predict and can be quite unreliable, even for “Gold” CD-ROMs.

Both of the projects performing tape backups (Case Studies #3 and #4) are also keeping their tapes off-site. Case Study #4 ships their tapes to an established vendor, while Case Study #3 came up with a more creative solution for cutting costs while still maintaining a higher level of data preservation. This project’s simple, but very clever, practice of exchanging tapes with another library in the area every week provides many of the benefits of off-site storage (namely, a disaster in a server room), with almost no additional cost. Other simple solutions like this should be developed and documented for the use of under-funded cultural heritage institutions.

This study was fortunate to find an example of the consequences of failing to create a comprehensive backup plan, in Case Study #1. While it can be difficult and annoying to be concerned with the preservation of data produced by other organizations, it is crucial for cultural heritage institutions who serve as coordination centers for digitization to do so nonetheless. In this case, avoiding responsibility for backups led to significantly more house being spent attempting to recover data from smaller institutions, in addition to some permanent data loss.

5. Systems staff

A significant issue across all of these studies was the presence of staff dedicated to computer systems work. By those who are fortunate enough to have funding for such a

position, the importance of this person was emphasized enthusiastically (Case Study #4). For those who are trying to fund such a position, the current lack was drearily described as a “real hole in our staff structure” (Case Study #3). Additionally, the technical sophistication in regards in digital preservation of Case Study #4 was no doubt influenced by, among other factors, the presence of a technically experienced employee who can devote a full schedule to such concerns.

The practical truth is often that such a person is simply unaffordable. Even in Case Study #4, where there is significant institutional support, the upcoming transition from external to internal funding has created some anxiety about the ability to continually fund a technical position. To support these cases, this study recommends:

1. Better best practice research and documentation from the field.
2. The creation of avenues for sharing innovative solutions to technical problems, such as the ingenious off-site tape storage scheme devised by Case Study #3.
3. The creation of a digital preservation certification for off-site repository facilities.

Conclusion.

The case studies suggest that many smaller cultural heritage institutions are unprepared for the transition of digital projects to digital programs. Some areas, such as the creation of archival masters, are generally strong. Other areas, like digital storage management, are often weak and are inconsistent across institutions. Additionally, the case studies generally lacked a strong plan for making the transition from external to

internal funding sources. Further studies are needed to more specifically identify the areas of primary need for institutions attempting the transition from project to program.

The tendency of sparsely-staffed institutions to outsource repository management seems to support the need for a digital preservation certification of repositories as suggested by Waters and Garrett (1996) and CLIR (2001). It can often be difficult to determine bit preservation activities of an off-site repository, especially for an institution without technical staff. By commissioning a certification of digital preservation soundness, a national institution would resolve a significant issue with repository management for cultural heritage data.

Finally, the study results suggest a need for better “best practice” research in the area of digital repository storage management. The similarity in the digitization specifications in the case studies show the promise of such research. The extensive publications, workshops and other activities surrounding the “best practices” for creating an archival master have clearly influenced the decisions of these smaller cultural heritage institutions. In contrast, the lack of activity in the area of digital repository management has led to widely divergent practices in the areas of digital storage media use for repositories and backups. These types of institutions may be more vulnerable to data loss as a result. Expanding the scope of digital preservation activities included in the field’s “knowledge environment” (CLIR, 2001) will help to alleviate these risks to cultural heritage records.

Bibliography

- Chapman, S. (2003). Counting the Costs of Digital Preservation: Is Repository Storage Affordable? *Journal of Digital Information (JoDI)*, 4(2). Retrieved 15 February 2006 from, <http://jodi.tamu.edu/Articles/v04/i02/Chapman/>
- CLIR - Council on Library and Information Resources. (2001). Building and sustaining digital collections: Models for libraries and museums. Washington, DC: Council on Library and Information Resources. Retrieved on 20 April 2006 from, <http://www.clir.org/pubs/abstract/pub100abst.html>
- Connaway, L. S., & Lawrence, S. R. (2003). Comparing library resource allocations for the paper and the digital library: An exploratory study. *D-Lib Magazine*, 9(12). Retrieved 20 February 2006 from, <http://www.dlib.org/dlib/december03/connaway/12connaway.html>
- Conway, P. (1996). *Preservation in the Digital World*. Washington, DC: Commission on Preservation and Access. Retrieved 12 November 2005 from, <http://www.clir.org/pubs/reports/conway2/>
- Eaton, F. (1994). Electronic Media and Preservation. *IASSIST Quarterly*. Spring/Summer 1994, 14-17.
- Greenstein, D., & Thorin, S. E. (2002). *The Digital Library: A Biography*. Washington, DC: Council on Library and Information Resources. Retrieved 10 April 2006 from, <http://www.clir.org/pubs/reports/pub109/contents.html>

- Greenstein, D., & Beagrie, N. (2000). A strategic framework for creating and preserving digital resources. A JISC/NPO study within the electronics libraries (eLib) Programme on the Preservation of Electronic Materials. *Multimedia Information and Technology*, 26(1), 84-85.
- Hedstrom, M. (1998). Digital Preservation: A Time Bomb for Digital Libraries. *Computers and the Humanities*. 31, 189-202.
- Kenney, A. R. (2000). Projects to Programs: Mainstreaming Digital Imaging Initiatives. In A. R. Kenney & O. Y. Rieger (Eds.). *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (pp. 153-175). Mountain View, CA: Research Libraries Group.
- LeFurgy, W. G. (2002). Levels of service for digital repositories. *D-Lib Magazine*, 8(5). Retrieved 25 March 2006 from, <http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html>
- Lesk, M. (1992). *Preservation of New Technology: A Report of the Technology Assessment Advisory Committee to the Commission on Preservation and Access*. Washington, DC: Commission on Preservation and Access. Retrieved 15 November 2005 from, <http://palimpsest.stanford.edu/byauth/lesk/lesk2.html>
- Lesk, M. (1996). Substituting Images for Books: The Economics for Libraries. Retrieved 20 April 2006 from, <http://www.lesk.com/mlesk/unlv/unlv.html>
- OCLC/RLG PREMIS Working Group. (2004). Implementing Preservation Repositories for Digital Materials: Current Practice and Emerging Trends in the Cultural Heritage Community. *Report by the joint OCLC/RLG Working Group Preservation Metadata: Implementation Strategies (PREMIS)*. Dublin, Ohio:

- OCLC Online Computer Library Center, Inc. Retrieved 25 March 2006 from,
<http://www.oclc.org/research/projects/pmwg/surveyreport.pdf>
- Rosenkrantz, G. J. (1971). National Archives Mass Storage Requirements—1975-1980. *IEEE Transactions on Magnetics*, 7(4), 843-847.
- Rothenberg, J. (1995). *Ensuring the Longevity of Digital Information*. Retrieved 10 November 2005 from, <http://www.clir.org/pubs/archives/ensuring.pdf>
- Shields, P. Overview: Backup media options. *MacTech Magazine*, 16(3). Retrieved 25 April 2006 from,
<http://www.mactech.com/articles/mactech/Vol.16/16.03/BackupManagement/index.html>
- Smith, A. (2001). *Strategies for Building Digitized Collections*. Washington, DC: Council on Library and Information Resources. Retrieved 1 April 2006 from,
<http://www.clir.org/PUBS/abstract/pub101abst.html>
- Waters, D. J. (1992). *Electronic Technologies and Preservation*. Washington, DC: Commission on Preservation and Access. Retrieved 20 April 2006 from,
<http://www.clir.org/pubs/reports/waters/waters2.html>
- Waters, D., & Garrett, J. (1996). Preserving digital information: Report of the task force on archiving digital information. Washington, DC: Council on Preservation and Access. New York: Research Libraries Group, Inc. Retrieved 20 April 2006 from,
<http://www.clir.org/pubs/abstract/pub63.html>

Appendix A. Phone Interview Protocol.

This is Jason Casden from the University of North Carolina at Chapel Hill, calling regarding your participation in a study dealing with digitization cost models. I appreciate your contribution to this study.

As we stated in the fact sheet that was sent to you, your participation in this survey is completely voluntary. This means that you do not have to participate in this survey unless you want to. Additionally, your confidentiality is guaranteed. Do you have any questions?

Do I have your permission to begin asking you questions?

I have about 12 to 21 questions, and we should be finished in under 40 minutes. You don't have to answer those questions if you don't want to. In fact you don't have to answer any question that you choose not to answer. And that is fine. We will just skip that question and go on to the next one. Are you ready to begin?

Questions:

- 1) What format were the original items used in this project?
- 2) Was this project grant-funded?
 - a. If so, how much was the grant allotment, and how long was the grant period?

- 3) What equipment and software did you use for this project?
- 4) Are your images stored in-house or is storage outsourced?
 - a. What is your estimated yearly cost for this storage?
- 5) How many staff members in your organization dedicate at least 5 hours a week to maintaining and supporting server-related equipment, software.
 - a. How many total hours per week are spent working with these systems?
 - b. Do you have any full-time staff dedicated to these activities?
- 6) Did you need to hire new systems staff to support this digitization project?
- 7) What is the approximate yearly cost, including staff time, to maintain your digital materials?
 - a. How do you fund these costs?
 - b. Was this more or less than anticipated at the start of the project?
- 8) Have you been able to consistently maintain Internet access to your digitized images?
 - a. If not, how long were the gaps?
- 9) Have you had any data loss?
 - a. If so, approximately, what percentage of your images were lost?
- 10) Do you back up your images?
 - a. Using what technology?
 - b. How frequently?
- 11) Are you using redundant storage technology, like RAID?
- 12) Do you feel that the administration of your institution made funding digital projects a priority?

We've reached the end of the questions. Thanks again for your participation. Do you have any questions before we end the interview?

Thank you, and have a nice day.