

Jennifer M. Mantooth. Digital Preservation Activities and Attitudes in American and Canadian Academic Libraries. A Master's paper for the M.S. in I.S. degree. April, 2009. 59 pages. Advisor: Christopher Lee.

Digital materials are highly used resources in academic libraries and absorb a significant cost in their creation. While digital materials greatly improve access to and usability of information, these materials must be actively maintained and preserved. Recent surveys, however, indicate that many libraries are not implementing adequate digital preservation techniques to insure their investments. This study surveyed American and Canadian institutions regarding their digital preservation processes and attitudes, with a specific focus on the comparison of processes and attitudes in institutions that have participated in a national digitization or digital preservation grant (grant receiving institutions) and non grant receiving institutions. The results suggest that grant receiving institutions are just as likely to implement migration, emulation, or digital archeology for digital image files as non-grant receiving institutions and that migration is the most widely implemented preservation strategy and is perceived as more effective than emulation or digital archeology by ARL institutions.

Headings:

Digital Preservation -- United States

Digital Preservation -- Canada

Academic Libraries -- United States

Academic Libraries -- Canada

Digital Preservation Activities and Attitudes in American and Canadian Academic  
Libraries

by  
Jennifer M. Mantooth

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  
Information Science.

Chapel Hill, North Carolina

April 2009

Approved by

---

Christopher Lee

## **Introduction**

As the human experience continues to become increasingly intertwined with technological tools, its documents likewise become increasingly dependent on the surrounding technology. Today's archivists are faced with the task of preserving digital objects and the related technology required to make sense of them. However, Hedstrom argues that it is not solely the archive's concern. The unique aspects of digital objects require early intervention for successful long-term preservation. Thus the traditional preservation roles held by archivists are creeping across professional boundaries. Government agencies, corporations, and research libraries all hold considerable amounts of digital information that will need to be preserved into the future and it is the organization's responsibility to ensure preservation (Hedstrom 3).

A discussion of the nature of digital objects helps to explain their increasing ubiquity in libraries as well as the unique challenges of digital preservation. Electronic documents can improve accessibility to heavily used resources, allow users to manipulate and annotate the information, as well as increase access points to underutilized information. Certainly, these benefits play a role in libraries' decisions to include digital objects in their holdings. However, electronic documents have unique preservation needs that cannot be fully met by the implementation of traditional preservation processes. Terry Cook writes that unlike their paper-based predecessors, digital documents are not eye-readable (Cook 301). Instead, digital objects are dependent on technology, which is constantly evolving, for interpretation. Digital preservation is further complicated by the

digital object's reliance on contextual information and documentation (Ross 13).

Today's libraries, however, are not facing digital preservation empty handed.

The archival community has discussed the challenge of long-term preservation of electronic documents through the development of content preservation strategies: migration, emulation, and digital archeology and has published numerous articles discussing the perceived benefits and limitations of each. Librarians and archivists have both been involved in the development of open access digital preservation tools such as PRONOM, an online file format registry and the Reference Model for an Open Archival Information System (OAIS), which provides a framework for long-term digital preservation and access.

Wrubel writes that "Electronic resources constitute an increasingly significant portion of library collections, both in usage and cost" (Wrubel 225). It seems then, that it is in the best interest of libraries and archives to implement adequate digital preservation policies to insure the continuing utility of their most valued resources, and justify the expense of creating or obtaining them. However, there seems to be a large spectrum of awareness and application of digital preservation techniques. While the Koninklijke Bibliotheek (KB), for example, is actively participating in a robust digital preservation strategy to preserve both content and access long term, there seems a large majority of institutions that are not actively engaged in or perhaps even concerned with digital preservation.

In an article published in 2006 attempting to inform institutions of the benefits and processes of digitization, A. Bhatnagar expresses awareness that obsolescence is a huge threat to digital materials and is careful to note the high financial cost of

digitization. However Bhatnagar fails to mention the cost of a long-term preservation strategy to keep these objects readable, and thereby insure the institutions initial investment in their digitization. In light of this research, it may come as little surprise then, that a 2005 survey of the digital preservation practices of Museums, Libraries and Archives in the United Kingdom revealed that while 23 out of 26 institutions surveyed acknowledged digital holdings (Simpson 15), only 6 had a digital preservation policy in place (Simpson 9).

A similar study was conducted in Australia in 2005. Pymm and Lloyd interviewed library staff at four state libraries and the National Library of Australia (NLA) on the acquisition, preservation, and accessibility of digital materials at the library. Pymm and Lloyd state that their purpose was “to gain a broad overview of their approach to handling their digital collections and of concerns they may have about trying to ensure long-term accessibility to such collections” (169). The researchers found that while the state libraries were aware that many of their digital holdings were now obsolete, “no library reported any program to review holdings and establish any migration/copying program while the technology may still be available” (Pymm 176). Additionally, all state libraries reported that they lacked the resources for digital preservation and were looking instead to the NLA to take the lead, or to partner with other libraries and archives to preserve digital materials. (Pymm 170 -171).

Findings of other surveys, however, provide evidence of progress. A survey conducted by Michele Cloonan and Shelby Sanett indicates that a significant portion of libraries and archives are actively involved in digital preservation activities. Cloonan and Sanett conducted a survey of thirteen institutions and archival projects in the US in

2002. Each of the institutions surveyed employed a member of the InterPARES Preservation Task Force. The study reports that 10 of 13 institutions surveyed (approximately 75%) were involved in a digital preservation strategy. While the survey conducted in the United Kingdom and Australia sampled institutions from the entire population of libraries, archives, and museums in the country, the American study sampled institutions from the International Research on Permanent Authentic Records Electronic Systems (InterPARES) project. Kyong-Ho Lee et al. write that the InterPARES project is “a multinational research initiative...collaborating to develop the theoretical knowledge and methodology required for the permanent preservation of authentic records” (101). The fact that all of the institutions surveyed in the American study were institutions with demonstrated commitment to digital curation may account for the high percentage of institutions involved in a digital preservation strategy.

It seems that many institutions are not actively engaged in digital preservation and a survey of the literature may provide insight into one of the reasons why. The literature does not universally agree on various aspects of the effectiveness of digital preservation techniques, such as the cost of migration and emulation. Archival institutions must continue to support current work processes as they investigate potential preservation strategies, the contradictory literature may be viewed as a hindrance to the individual looking for a quick answer. It seems that institutions actively engaging in grant projects such as interPARES are more likely to display digital preservation activity. The realization of this implication appears to exist as a void in the current literature.

**Purpose:**

This research surveyed American and Canadian research institutions regarding their digital preservation processes and attitudes. The main focus of the study was to explore preservation activities and perspectives at these institutions. An additional focus was the comparison of processes and attitudes in institutions that have participated in a national digitization or digital preservation grant (grant receiving institutions) and non grant receiving institutions. The study also compared the attitudes towards digital preservation techniques with the literature.

**Research Questions:**

1. What are the general trends in digital preservation practices and attitudes at American and Canadian research institutions?
2. Are grant receiving institutions more likely to be actively involved in digital preservation activities?
3. Is there a relationship between the presence of digital preservation activities at the institution and the attitudes about the techniques used?

**Literature Review**

Overall, the literature offers valuable, high level discussions of digital preservation techniques, what perhaps is missing is an appreciation of the individual archival institution's perceptions of the preservation options available.

Much of the discussion on digital preservation techniques is published in online trade journals such as *D-Lib Magazine* and *RLG Digi-News* (now defunct). When possible, case studies are used to illustrate the theoretical benefits and/or perceived shortcomings of the approach.

### Reliance on Standards

Many digital preservation techniques, including migration and emulation combine reliance on standards with additional preservation activities. Rothenberg writes that “converting digital documents into standard forms, and migrating to new standards if necessary, may be a useful interim approach while a true long-term solution is being developed”. The Library of Congress defends the importance of formal technical standards for the successful creation and maintenance of digital objects, and especially digital images, as evidenced by a compendium of digital image standards published by the Library of Congress in 2004 (Peterson). Rothenberg, however, cautions that one of the limitations of relying on standards is that formats “are still evolving so rapidly that it is unrealistic to expect definitive standards for any of these forms to emerge in the near future” (Rothenberg). Yet even today, the literature indicates that formats continue to evolve and cloud the standards issue, as illustrated by the debate over the defacto standard for digital image archiving.

In 2008, Paolo Buonora and Franco Liberati write that digital libraries prefer JPEG 2000 to TIFF, and that this format has in fact replaced TIFF as the defacto standard for digital image archiving. Yet, while supporting JPEG 2000 as a viable preservation format, Robert Buckley cautions that “JPEG 2000 is lossy and irreversible, [although] the losses are not visually noticeable” (26). Andreas Stanescu writes that lossy compression is “an impediment to preservation” and advises the use of the TIFF file format over any lossy format (74). The controversy over the appropriateness of JPEG 2000 to replace TIFF as the defacto standard for digital images offers evidence that Rothenberg’s claim, issued nearly a decade ago, still holds true.



Rothenberg indicates that another limitation of reliance on standards as a digital preservation technique stems from the issue of forcing incompatible document types, or paradigms, into the 'standardized format'. Rothenberg uses the term paradigm to indicate the vastly different methods of display, information storage, and functionality of technical classes of documents such as databases, spreadsheets, e-mail attachments, and images. Rothenberg writes that reliance on standards necessitates paradigm shifts which may lead to a great loss of information as the original format is forced into a standardized form. Rothenberg indicates that when paradigms are drastically different, such as relational databases and object oriented databases, the translation of one to another loses so much information that the result risks becoming meaningless.

#### Digital Archeology

Computer museums are the critical force behind digital archeology. Digital archeology, also known as retrospective archiving (retro-archiving), refers to the process of retrieving information from obsolete media. One of the benefits of digital archeology is that the artifact can stay in its original state or paradigm and maintain its original functionality (Rothenburg). Digital archeology is not considered a responsible preservation technique on its own, and is generally combined with either migration or conversion to a standardized format. A limitation of this approach is the expense. Rothenberg writes that digital archeology becomes highly labor-intensive because the computer museums need to maintain hardware and software drivers to enable the obsolete machines to read the new storage media devices. Kyong-Ho Lee et al., add that computer museums and by extension digital archeology may be impossible to support over the long term due to equipment deterioration and lack of vendor support (95).

Despite their limitations, a study conducted in 2007 indicates that digital archeology is becoming an increasingly important preservation technique. In 2007, Balkestein and Tjalsma report the results of the Archiving Digital Academic Heritage project, a three-year project focused on using digital archeology to provide access to digital research datasets on obsolete media at the Meertens Institute in the Netherlands (95). Even though the data sets were on media that was ten to thirty years old, with the help of the Computer Museum of the University of Amsterdam, the project was able to read approximately 97% of the obsolete formats, only 52 of 1599 were unreadable (Balkestein and Tjalsma 96). While the project was technically successful, Balkestein and Tjalsma indicate that one of the greatest limitations of digital archeology is the loss of contextual information necessary to give meaning to the raw data. Without this contextual information the recovered files had much less value (96). The literature indicates that libraries and archives will continue to be heavily dependent on incorporating digital archeology into their digital preservation activities. An institutional report from the National Library of Australia suggests that without the external help of a computer museum approximately 1/3 of obsolete media (in this case up to 12 years old) will be unreadable (Colin).

### Migration

In *Network Notes #60*, published by the National Library of Canada, Alison Bullock defines migration as a process that “covers a range of activities to periodically copy, convert or transfer digital information from one generation of technology to subsequent ones.” Oltmans indicates that a benefit of migration is that the objects will usually be in a format that is easily accessible across various computing platforms.

Rothenberg offers a scathing review of migration, which reinforces his role as what Granger deems “the staunchest supporter of emulation”. Rothenberg lists among the limitations of migration that it is error-prone, expensive, and difficult to automate.

Much of the literature agrees that one of the largest weaknesses of migration is the potential for data loss. Oltmans cautions that a single conversion from one form to another may result in loss of data. Mellor discusses the potential for a process of compounding migration errors from one migration to another throughout future migrations and resulting in an object that may have little resemblance to the original (517). Bullock agrees that successive migrations may have the potential to lead to unacceptable data loss.

Rothenberg indicates that much of the expense of migration resides in the inability to safely automate the process. He writes that each different document type requires a unique solution, independent technology forecasting to determine when migration is necessary, and variable migration schedules, all of which require independent handling and cannot be responsibly applied to different document types (Rothenberg). Mellor attributes the high cost of migration to the need to continually produce migration tools and then apply them to a very large number of data objects. Granger argues against the attribution of ‘expensive’ as a shortcoming of migration, he asks “expensive compared with what?” and writes that migration has been the “only serious candidate thus far for preservation of large scale archives.”

Despite the perceived shortcomings of migration, Kranch, Bullock, Granger, and Rothenberg agree that migration is the most common digital preservation strategy among archival institutions. The OAIS Reference Model cautions that migrating digital

content is time consuming, expensive, and may expose the repository to a “greatly increased” probability of loss of information (Reference 91) Despite the caution, the OAIS model seems to prefer migration over emulation techniques which the authors write are not mature enough for “significant comment” (101) The model contains a detailed list of potential motivations for migration (90 – 91), offers a migration perspective for the model (92) and discusses various migration types (93 – 98).

In an effort to allay the costs associated with constant migrations to the continually evolving popular formats, Mellor proposes a variation on traditional migration: migration on request. Unlike traditional migration, which hinges on processes that transform the copy of data that is stored within the repository, migration on request preserves the original digital object and migrates the object, using an algorithm based on its format, only when the object is requested for viewing (Mellor 518). Based on the results of a migration on request tool, developed by Mellor and a team of researchers, Mellor indicates that “over a short to medium term period migration on request should offer major cost savings in comparison to a traditional migration strategy” (526).

### Emulation

While there are significant amounts of literature dealing with emulation, the literature seems to confuse an already technically complex issue and there is argument over the perceived benefits and limitations of emulation. This confusion may offer insight into the overwhelming preference for migration in archival institutions, as well as the tendency of archival institutions with demonstrated commitment to digital

curation, such as participation in the InterPARES programs, to be more engaged in digital preservation.

Oltmans defines emulation as a process which “preserves the authentic document and provides the user with a tool that enables ‘old’ software and ‘old’ viewer programs to render this original document”. In Oltman’s definition of emulation, the document and the application software are original, and the emulation imitates either the operating system or the underlying hardware. However, Bullock defines emulation as the process of “creating new software that mimics the operations of older hardware of software” While Bullock’s definition may be more encompassing than Oltmans (she includes the potential for all three types of emulation: software, operating system, and hardware), it indicates the difficulty of making sense of the literature. One may suspect that each type of emulation carries its own challenges and positive affordances. However, when emulation is dealt with in the literature, it generally is not prefaced by an identification of type.

Another confusing factor is the issue of the ability to test an emulator beforehand. Again, the literature is not consistent with respect to this aspect. While Rothenberg writes that one of the great strengths of emulation is the ability to test the accuracy of the emulator in advance, Oltmans indicates that one of the weaknesses of emulation is that one cannot know for certain that they will always work on future computer platforms. The reader may assume that the difference may depend on the type of emulation and the length of time projected, but it still does little to illuminate the issue.

One thing that the literature consistently agrees upon is the ability for emulation to preserve the original ‘look and feel’ of the digital object. (Oltmans, Bullock, Rothenberg, and Granger). However Margaret Hedstrom and other researchers conducted an experiment in 2006 in which thirty subjects evaluated the performance of a twenty-year-old computer game in its native environment, in an emulated environment, and migrated to a Windows environment (168). The team reports that they “found no evidence that emulation is better for preserving original ‘look and feel’ than migration” (Hedstrom et al. 171). In fact, the study reports that few subjects actually lamented the loss of the original ‘look and feel’ of the computer game, most instead preferring the improved functionality of the migrated and emulated versions (Hedstrom et al. 171). It may be interesting to point out that in this experiment the display of the original was not copied exactly in the emulated version which imposed a bar across the screen (Hedstrom et al. 170).

A cost comparison of migration and emulation by Erik Oltmans and Nanda Kol of the Koninklijke Bibliotheek (KB) in the Netherlands provides contrary evidence to the generally accepted claim that emulation is more expensive than migration. The KB has collaborated with IBM to develop and implement the e-Depot, a system that preserves both the content of and accessibility to digital publications (Oltmans and Kol). The validity of the research team’s findings may be supported by the evidence that the KB is acclaimed worldwide for its expertise in digital preservation and in fact has gained the confidence of 6 major publishers including Elsevier Science and Springer to archive their electronic publications (Oltmans and Kol). Generating cost

comparisons from literature and personal insights, Oltmans and Kol conclude that, in the long term, emulation is more cost efficient than migration.

Oltmans and Kol base their conclusion on the issue that migration entails a steady stream of resources to continually monitor and convert objects; emulation on the other hand requires high initial costs (Oltmans and Kol). A significant virtue of emulation, in the realm of costs, is that once it is created an emulation tool can be applied to an entire collection of objects as opposed to a repetitive single object at a time approach as in migration (Oltmans and Kol). Oltmans and Kol argue that it is this property of emulation that makes it more cost effective than migration for long-term preservation. According to a cost projection for preserving 1 million objects for 50 years, Oltman and Kol report that the cost of migration is 79% more than the cost of emulation. When the authors increase the number of objects in the archive to 5 million, after 50 years the projected cost of migration is more than 200% over the cost of emulation. Although perhaps the bulk of the costs for emulation reside in the creation of the emulator itself, Oltmans and Kol advise that these costs can be distributed and thereby lessened if digital repositories globally share their emulators with each other.

#### Other Techniques:

While much of the literature discusses well-accepted digital preservation techniques, the literature does contain processes that have not made it into the mainstream. Reliance on print outs and digital tablets are examples of such techniques.

#### Reliance on Print Outs

Jeff Rothenberg defines reliance on hard copy as to print digital documents onto paper and then follow traditional archival processes to preserve that paper. A benefit to

this technique would be that archives and libraries could use traditional physical preservation techniques that have been used to preserve paper based archival holdings. However, the translation of a digital born object to an analog document has the potential to lose many if not most of the positive affordances innate to that object, and also a potentially great deal of information as well. Margaret Hedstrom and Christopher Lee define significant properties as “those properties of digital objects that affect their quality, usability, rendering, and behavior” (218). Rothenberg notes that a significant limitation of reliance on print outs is that it forfeits the benefits of machine readability and obliterates the interactive functionality within and between digital documents.

#### Digital Tablets

In 1998, Douglas Kranch published an article that champions digital tablets as a superior and inevitable alternative to migration and emulation. Kranch’s digital tablets would preserve the hardware and software required to access the digital objects, as well as the object itself. In addition, the tablets would contain a large amount of read only storage (10s or hundreds of terabytes), reside in a protective plastic encasing, and run off their own power supply. Kranch indicates that a benefit to digital tables would be the ability to preserve the original document and its software in order to preserve the functionality and “feel” of the document. Despite the initial attention, digital tablets are not mentioned as a viable preservation strategy in later literature.

#### Stepping back

While most of the literature agrees that reliance on print outs, reliance on standards, and digital archeology are not valid preservation strategies on their own, there continues to be discussion over the comparative strengths and weaknesses of



migration and emulation. Much of the literature advises the individual to step back from the quest for the “silver bullet” preservation strategy and examine the issue as a whole.

#### Contributing factors

While technological strategies no doubt have an impact on the success of a preservation program, the literature suggests that other factors come into play as well. Rothenberg indicates that although “a sound technical approach must form the foundation on which everything else rests”, administrative, organizational, and policy issues also impact the success of a preservation strategy. In contrast to Rothenberg, Nancy McGovern advocates that digital preservation be supported by three main components, which can be seen as three legs of a stool: organization, technology, and resources. She advises that “a balanced three-legged stool ... will be more effective in establishing a sustainable digital preservation program than a technology pogo stick” (McGovern).

A case study conducted by Howard Besser in 2006 attributes the success of the Preserving Digital Public Television project, a highly collaborative digital preservation project engaging 3 large, high profile public television stations and New York University, in part to the organizational commitment of the parties involved. Besser writes “from early on in the project there was active high-level commitment and support from most of the players (at the vice-president level from the two stations, at the dean’s level from the university).” (223).

#### Dual approach

While authors exist on either end of the spectrum, much of the literature advocates a combination of digital preservation techniques. Both Rothenberg and

Bullock support the use of standards in maintaining cost efficiency. Mellor advises institutions to implement migration on request along with emulation (526).

Different approaches support different needs

The literature encourages the archival institution to understand the unique characteristics of their own collections when choosing preservation techniques. Oltmans and Kol discuss the importance of what Hedstrom and Lee term as the significant properties of a digital document when selecting a preservation strategy. If an organization has a commitment to preserve the authentic form of a document, or if the original look and feel is deemed a significant property of the object, Oltmans and Kol propose emulation over migration. However, if an archive anticipates that future users will benefit from viewing an object according to the standards and functionalities they have come to expect from technology migration is a better strategy (Oltmans and Kol).

### **Methodology**

This research gathered information about the digital preservation activities of academic libraries and the perspectives of key members of these institutions regarding well-known digital preservation techniques. Academic libraries will be the units of analysis for this study. In order to study the practices and attitudes of institutions, this study surveyed individuals who acted as informants for the institutions. Surveys are considered an effective tool for measuring attitudes (Babbie 244). Informants provided information regarding the digital preservation practices of the institutions and the institutional perspective on digital preservation techniques. Informants were recruited through email to participate in a web-based study. This enabled the study to examine institutions without the limitations of geography. The survey targeted individuals with

a fair amount of authority in the institution, based on the assumption that they are aware of digital preservation activities taking place at the institution. Additionally, the digital preservation perspectives of these individuals may have played a larger role in the institutional perspective. Most of the survey questions were close-ended, enabling respondents to choose an answer from a list. A benefit to close-ended questions is that they “provide uniformity of responses and are more easily processed than open-ended questions” (Babbie 246). Care was taken to ensure that the survey questions were concise and addressed only a single topic.

#### Sample, Population, and Sampling Technique

This study recruited contacts from the Association of Research Libraries (ARL) member list from the ARL website. ARL is a not-for-profit organization of research libraries in the United States and Canada. There are currently 123 members of the ARL.

#### Subject Identification and Recruitment:

All the academic libraries surveyed came from a variety of American and Canadian institutions with digital collections. 17 of the 26 respondents, who indicated their participation in grants, had received a national digitization and/or digital preservation grant, 7 had not received one, and 2 did not know.

Libraries do not generally follow the same organizational hierarchy, and job roles vary widely across institutions. However, a list of head administrators of each library and their emails was constructed by the author. These individuals were emailed a request to forward the survey to the individual or individuals at their institution who are most capable of providing answers regarding the institution’s digital preservation

practices. One week after the initial email was sent, a follow up email was sent to encourage response.

### Variables:

Previous participation in a national digitization or digital preservation grant was a quasi-independent variable in this study. Digital preservation practices and attitudes was analyzed independently within these groups and then against each other.

Engagement in migration, emulation, and digital archeology also acted as a quasi-independent variable in later data analysis.

### Data Collection Instrument

Data was collected by means of an online survey. The survey was divided into three parts. The first part provided identifying characteristics of the informant and the institution. The following five questions composed the first part of the survey:

1. *Please identify the location of your institution (U.S. Canada Other)*
2. *Please indicate your job title.*
3. *Please provide the name of your institution. (Note: this information is used to identify duplicate responses and will not be included in any derived report).*
4. *Within the past 10 years, has the institution been involved in a national digitization and/or digital preservation grant?  
(Y N Applied but didn't receive funding Don't know)*

Institutional names were replaced with a generic identifier (A,A1, etc) at the close of the survey in order to maintain the confidentiality of institutions in the raw data and in the analysis.

The fifth question asked respondents to describe the contents and size of their institution's digital collections.

The second part of the survey provided information regarding the institutions digital preservation activities.

In the following two main and sub questions, the informant was asked to provide information regarding their digital text and image files. These questions were open-ended to encourage honesty. The large number of potential file formats precluded the use of a list of possible answers, and at any rate such a list arguably would have never included all possible formats. These questions provided insight into the institution's use of file standards in their preservation activities. Additionally, they helped indicate if the institution was following best practice regarding storage media and file refreshing timelines. The final function of these questions was to verify the truthfulness of the next question. That is, if an informant responded that the preservation copies of their image files were mostly on the original 15 year old CDs and later indicated that the image files had been migrated, the researcher could note the inconsistency.

5. *Please provide the following information for digital TEXT files:*
  - a. *What file format(s) do you use for the digital master copies of your digital text files?*
  - b. *On what physical media do you store your digital text collections?*
  - c. *Are you storing your digital text collections on physical media that are different from those to which you originally stored them?*
  - d. *What is the age range of these materials?*
6. *Please provide the following information for digital IMAGE files:*

- a. *What file format(s) do you use for the digital master copies of your digital image files?*
- b. *On what physical media do you store your digital image collections?*
- c. *Are you storing your digital text collections on physical media that are different from those to which you originally stored them?*
- d. *What is the age range of these materials?*

The next questions were arranged in a table to enable ease of response. Survey participants were asked to indicate institutional participation in migration of image files and of text files, emulation of image files and of text files, and digital archeology of image files and of text files. Possible answers were Yes, No, or Don't Know. Text and image files were handled separately in order to avoid barrel questions. Digital archaeology was defined in the questions as 'extracting information from obsolete media – yourself or by hiring a vendor to do the extraction' in order to clarify the technique to participants.

7. *Please indicate whether your institution has used migration in the management of your digital image files*
8. *Please indicate whether your institution has used emulation in the management of your digital image files*
9. *Please indicate whether your institution has used data archeology (extracting information from obsolete media – yourself or by hiring a vendor to do the extraction) in the management of your digital image files*

10. *Please indicate whether your institution has used migration in the management of your digital text files*
11. *Please indicate whether your institution has used emulation in the management of your digital text files*
12. *Please indicate whether your institution has used data archeology(extracting information from obsolete media – yourself or by hiring a vendor to do the extraction) in the management of your digital text files*

The third part of the survey provided information regarding the informant's perspective on migration, emulation, and digital archeology. The informant was asked to provide his or her personal reaction to the statements that followed. As stated earlier, informants were selected based on their knowledge of digital preservation practices in the institution. Even though the institution is the unit of analysis for the study, in this case, the informant was the closest voice of the institution available. Responses to these questions were based on a five-point Likert scale (Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, or Strongly Disagree). This format is generally quick to complete, and can increase the ability to compare responses across questions (Babbie 254). All five statements about a given preservation technique as well as the corresponding Likert scale were grouped to be visible on one computer screen for each preservation technique. These questions provided comparable information about the attitudes of the informant regarding three major digital preservation techniques and enabled the researcher to create descriptive statistics of digital preservation

perspectives. The instruction to the informant and the statements for each technique are listed below.

*Please indicate your level of agreement with the following statements:*

*Migration:*

*is labor intensive*

*has a high financial cost*

*is an effective preservation strategy for digital text*

*is an effective preservation strategy for digital images*

*is widely implemented*

*Emulation:*

*is labor intensive*

*has a high financial cost*

*is an effective preservation strategy for digital text*

*is an effective preservation strategy for digital images*

*is widely implemented*

*Digital Archeology (extracting information from obsolete media – yourself or by hiring a vendor to do the extraction):*

*is labor intensive*

*has a high financial cost*

*is an effective preservation strategy for digital text*

*is an effective preservation strategy for digital images*

*is widely implemented*

Advantages and Disadvantages of Method



Babbie indicates that the strengths of surveys include the ability to describe characteristics in a large population and the ease of information gathering from a large sample size (276). The weaknesses of surveys include the inability to delve deeply into complex issues in the same manner that observational methods can and the inability to modify the tool after the study is underway (Babbie 276).

## **Results**

Of the 123 ARL institutions, the researcher was only able to discover 113 top library administrators and their email addresses. 113 heads of the libraries at ARL institutions were recruited to participate in the survey and 43 surveys were started. Surveys in which the respondent exited immediately after clicking “accept” on the consent form, which was the first question of the survey, were removed from the resulting dataset. This removed 7 surveys from the resulting dataset, and indicates that the true response rate is 34/113 or approximately 30%.

### Duplicate responses

The survey respondents are informants into an ARL institution’s digital preservation activities and their own attitudes. Due to the nature of survey distribution, the researcher anticipated that she may receive several responses from different informants at the institution. However, the unit of analysis for this study is the ARL institution, so it was deemed important that no institution be over-represented in the resulting dataset. In order to ensure that one institution was not overly represented in the resulting data, the survey results were examined in order to identify duplicates. IP addresses, institution names, and the job titles of respondents were examined and 5 institutions appeared to have duplicate survey responses.

Three institutions had two surveys taken by the same individual. Based on the IP address, the institution name, the respondent's job title, and the respondent's note in one of the partial surveys, the researcher was able to combine two partial surveys taken by the same individual at one institution. The researcher was contacted by one participant who completed two surveys for the same institution and was informed that the most recent survey was most accurate, as a result she removed the older duplicate from the resulting dataset. In the same manner and in order to minimize the effects of duplication, the less complete survey for the third institution was removed.

The fourth and fifth institutions both had two surveys taken by two different individuals. In both cases, one survey was minimally complete and the other was complete. The responses were identical except for the job title. As a result, the researcher removed the partially complete survey from the resulting dataset. The resulting, non duplicate dataset consists of 31 surveys.

#### Respondent background

28 known institutions were represented in the 31 responding, non duplicate surveys. 3 respondents did not identify the name of their institution. Of the 31 institutions represented, 27 were in the United States and 4 were in Canada.

Of the 31 survey responses, 3 respondents did not identify their job title. Of the 28 respondents that provided their job title, 20 (over 70%) held positions of considerable authority in the library. Three respondents identified themselves as the head of the library (Director of Libraries, University Librarian, Dean of Libraries), one was the Deputy Director, and 16 respondents held job titles that indicate that the respondent was the director or head of a department within the library.

### Specialization

Many surveys were completed by individuals that held job titles that indicate specialization in digital technologies. 18 of the 28 respondents had the term “digital” or “technology” in their job titles. 6 of 28 respondents had the term “preservation” in their job title. Four of the 6 respondents with the term “preservation” in their job title also contained the term “digital” Only 2 individuals surveyed held job titles that contained any variation on the term “archivist.”

### Participation in Grants

Five respondents did not answer this question. 17 of 26 respondents (65%) indicated that their institution had been involved in a national digitization and/or digital preservation grant. Two of 26 (8%) indicated that a grant had been applied for but not funded, and five of the 26 respondents (19%) for this question indicated that their institution has not been involved in a national digitization and/or digital preservation grant. Two respondents did not know if the institution had been involved in a digitization grant in the past 10 years.

### Digital Collections

Respondents were asked to describe the contents and size of their institution’s digital collections. 25 informants responded to this question. Fourteen respondents indicated the size of the digital collections only in terms of number of items. Five respondents reported the size only in terms of Terabytes, Gigabytes, or Megabytes. Four gave both item counts and approximate number of bytes to store the materials. Two provided no indication of size in either number of items or number of bytes.

The number of items reported in an institution's digital collections varied in precision from '11,688' to 'several tens of thousands.' They varied in size from one institution that reported 'approximately 2,200 images and one serial title' to approx. 2,405,847 text, images, video, and learning objects. Of the 15 respondents who provided a specific number of items, the mean number of items reported is 233,080. The median number of items reported is 22,147.

Nine respondents provided the size of their digital collections in terms of number of bytes. The answers ranged from 52 gigabytes to 'no smaller than 40 terabytes'. The mean size of digital collections reported is 15 terabytes. The median is 5 terabytes.

Three respondents indicated that they anticipate their digital collections to increase in size in the near future.

Collection contents varied. 24 of the 25 respondents indicated some combination of text, image, audio, and video collections. One respondent provided no indication of their collection contents. Five respondents indicate that their digital collections include theses, dissertations, and scholarly publishing texts. Several respondents list websites (2) and statistical datasets (4) among their digital holdings.

### Digital Text Files

Respondents were asked to indicate the file formats used for the digital master copies of their digital text files. There were 25 responses listing various file formats. Among these responses there is a surprising presence of image based file formats. This indicates that respondents were including scanned page images in this category, rather than limiting their responses to born-digital text files. The text file formats mentioned

by respondents include PDF, TXT/ASCII and RTF. Figure 1 illustrates the distribution of file formats in the responses.

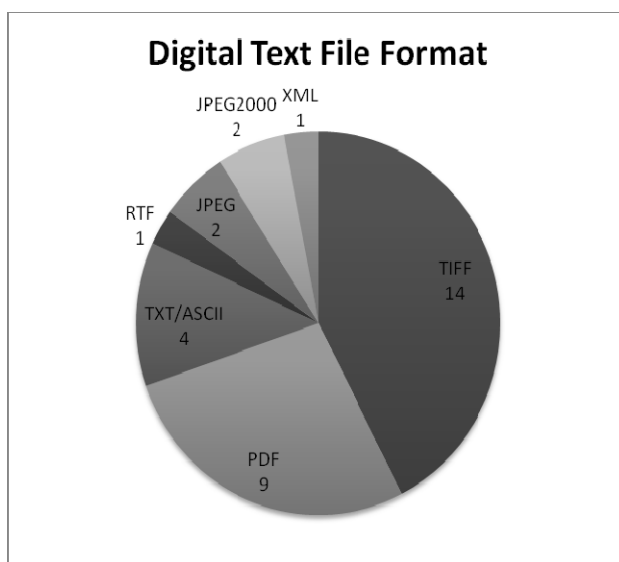


Figure 1: Distribution of File Formats for Digital Text Files

All respondents reported that their digital text collections are stored on some variation of what may be labeled large networked mass storage devices. Four respondents indicate that CDs and/or DVDs are used in addition to servers. One institution reports that while they do store some items on a server, most of their purchased collections are housed and accessed from outside of the institution. 13 informants indicate that their institutions are storing their digital text collections on physical media that is different from the original storage media. Three report that the media is not different. 2 indicate that some media has been refreshed. 1 did not answer. 2 did not answer the question directly but indicated that backup copies are on DVDs or CD-ROMs. Three did not know or wrote N/A.

The age range of the digital files varied from less than 1 year to up to 20 years old. The distribution of files by age can be viewed in figure 2.

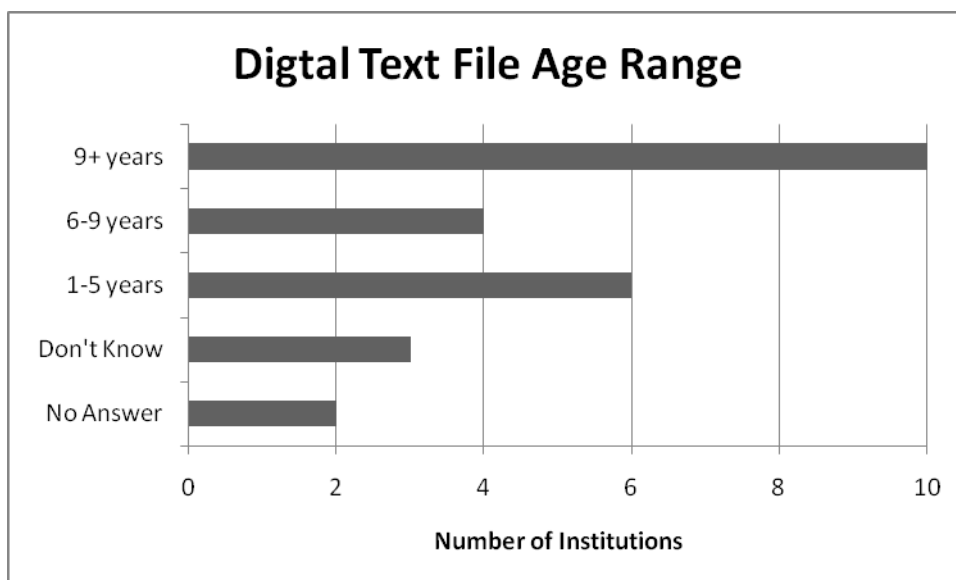


Figure 2: Distribution of the age range of digital text files at surveyed institutions.

### Digital Image Files

Respondents were asked to identify the file formats used for the digital master copies of their digital image files. Most respondents indicated formats for single images, while one included file formats for documents that may comprise of scanned text with embedded images (PDF, PDF/A). There were 26 responses, some of which listed a variety of formats. 20 responses indicated only TIFF. Three responses listed TIFF along with JPEG and/or JPEG2000. One response listed TIFF, PDF, and PDF/A. Figure 3 illustrates the distribution of file formats in the responses.

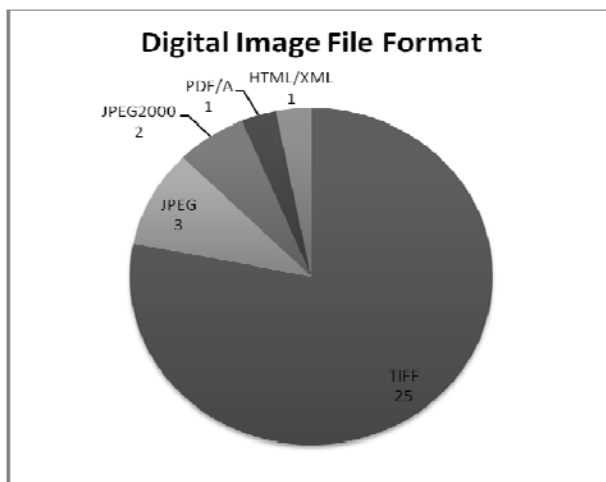


Figure 3: Distribution of File Formats for Digital Image Files

The physical media used to store digital image collections match with the physical media reported to store digital text collections (large networked mass storage devices) for 25 of the 26 institutions that responded to this question. One respondent indicates that while all the digital text files at his/her institution resides on an ‘archival server’, some image files are on CD and the rest are stored on an ‘archival server’. The respondent did not provide clarification on the term ‘archival server’.

The distribution of institutions that store their digital image collections on media that is different from the original is similar to the counts for digital text collections. 15 respondents indicated yes. 5 indicated no. 3 indicated that some media has been refreshed. 2 did not know or wrote N/A.

The age range of the digital files varied from less than 1 year to up to 20 years old. The distribution of files by age can be viewed in figure 4.

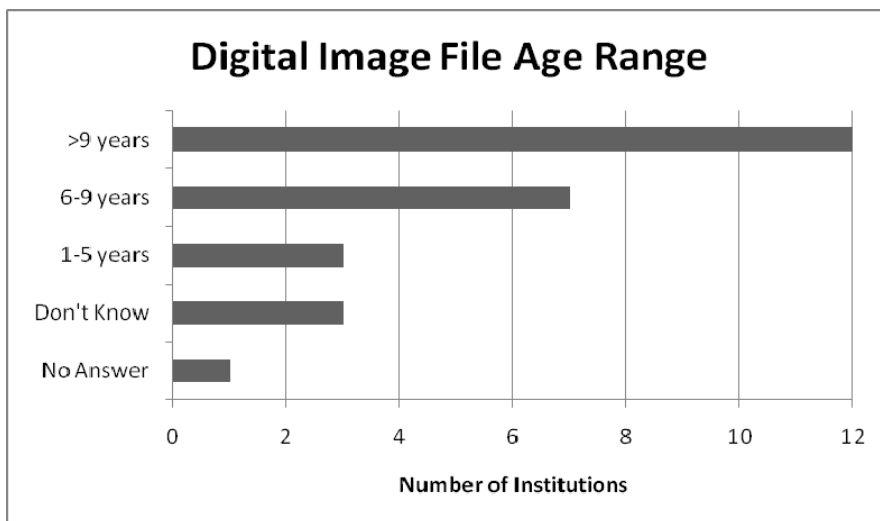


Figure 4: Distribution of the age range of digital image files at surveyed institutions.

### Digital Preservation Practices at ARL Institutions

Respondents were asked to indicate their institution's use of three digital preservation techniques (Migration, Emulation, and Data Archaeology) in the management of its digital image and text files.

24 individuals provided information regarding the institution's use of migration. 25 respondents answered the question about the institution's use of emulation and data archeology.

Participation in migration, emulation, and data archeology was sub-grouped based on institutions that have received a grant in the past 10 years and those that have not. The percentage responses are similar for both groups with slight differences. Statistical tests were done to determine the statistical significance of these results. There was no significant relationship found between receiving or not receiving a grant and participating in migration or emulation of image files ( $\chi^2 = .8$ ,  $df = 1$ ,  $p = .37$ ). There was no significant relationship found between receiving or not receiving a grant and participating in digital archeology or migration of image files, ( $\chi^2 = .02$ ,  $df = 1$ ,  $p = .88$ ).



And there was no significant relationship found between receiving or not receiving a grant and participating in digital archeology or emulation of digital image files ( $\chi^2 = .89$ ,  $df = 1$ ,  $p = .35$ ). There was not enough data for text files to perform statistical analysis.

Therefore, statistically, non-grant institutions are just as likely to participate in migration, emulation, or digital archeology for their digital image files as grant institutions. The percentage differences viewed in figure 5, below, are not statistically significant.

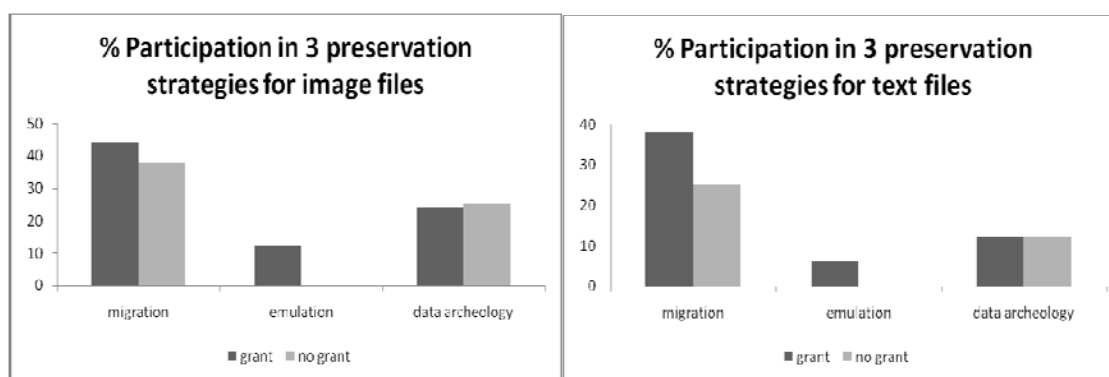


Figure 5: Percent participation in migration, emulation, and data archeology for grant and non-grant receiving institutions. The chart on the left shows the percentages for image files, the chart on the right shows percentages for text files.

The standard deviation for grant receiving institutions that participated in migration for their digital image and text files is .62 and .68, respectively. The standard deviation for non-grant receiving institutions that participated in migration for their digital image and text files is .71 and .64 respectively. The standard deviation for grant receiving institutions that participated in emulation for their digital image and text files is .50 and .49, respectively. The standard deviation for non-grant receiving institutions that participated in emulation for their digital image and text files is 0 and .35 respectively. The standard deviation for grant receiving institutions that participated in

digital archeology for their digital image and text files is .53 and .50, respectively. The standard deviation for non-grant receiving institutions that participated in digital archeology for their digital image and text files is .64 and .53 respectively.

Due to the strong similarities between the use of migration, emulation, and data archeology between grant receiving institutions and non-grant receiving institutions, the results of the survey will be discussed in terms of the overall findings, as both groups of institutions follow the same trends as the group to which both belong.

As illustrated in figure 6, more institutions report using migration, emulation, and data archeology on their digital image files than on their text files. Additionally, more institutions are using migration than either data archeology or emulation.

Emulation is the least used preservation technique.

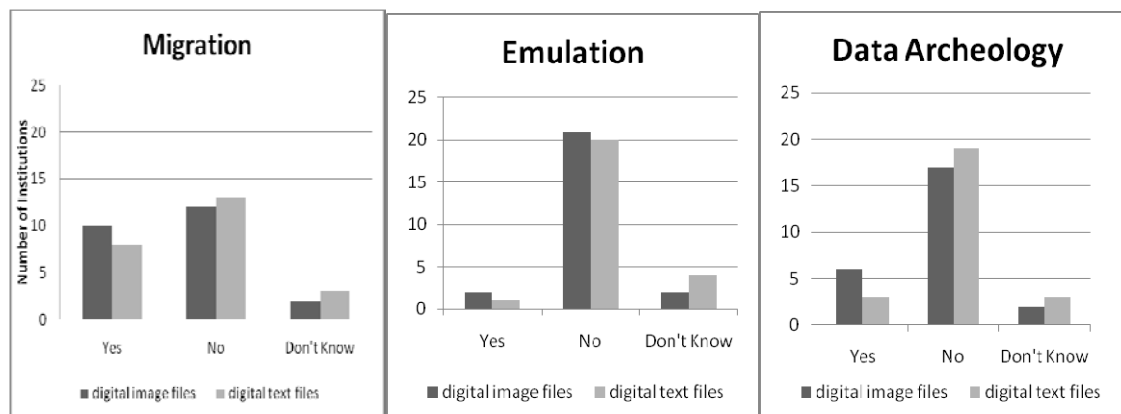


Figure 6: Distribution of the use of migration, emulation, and data archeology

Figure 7 illustrates the relationship between the counts of institutions engaged in migration, emulation, and data archeology of digital image files and the counts for digital text files. Approximately 42% of institutions surveyed have used migration in the management of their digital images and 33% have migrated digital text. 24% of respondents indicated that their institutions have used data archeology on their digital image files, while 12% have used data archeology on digital text files. The surveys

indicate that much fewer institutions are engaged in emulation. 8% of respondents indicate that their institutions have used emulation for digital images and 4% indicate that they have used emulation for digital text.

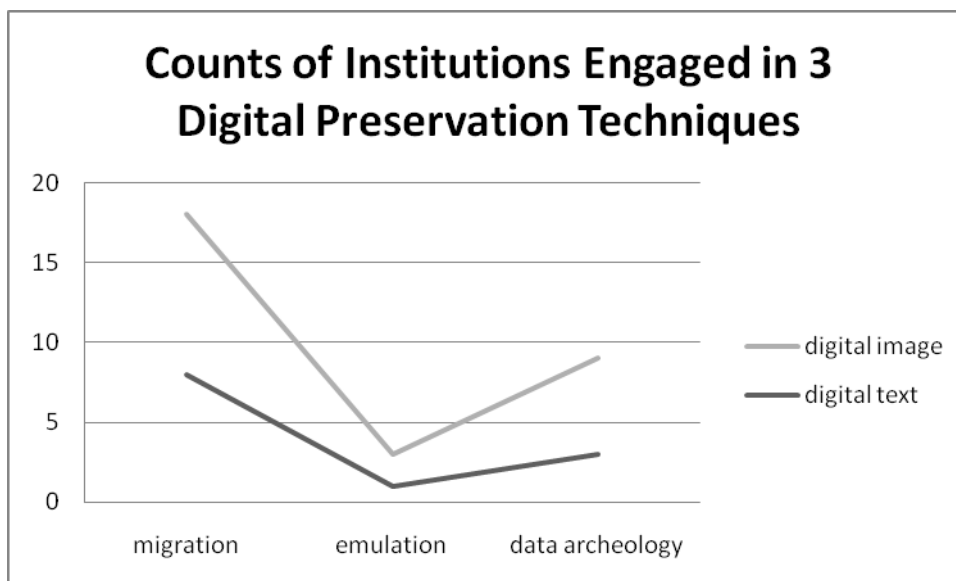


Figure 7: Counts of Institutions engaged in migration, emulation, and data archeology for their digital images and digital text files

In addition, institutions are more likely not to use migration, emulation, or digital archeology than to use them, regardless of file type. Figure 8 illustrates that all counts of institutions using migration, emulation, or data archeology are lower than the count of institutions that are not engaged in them.

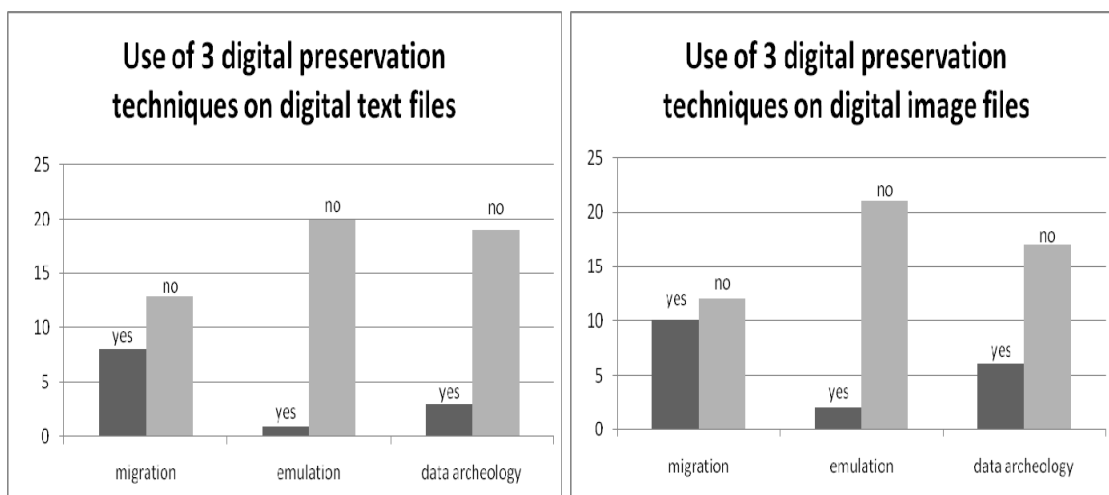


Figure 8: Use of digital preservation techniques by type of file.

### Digital Preservation Perceptions at ARL Institutions

Respondents were asked to indicate their level of agreement with statements issued about different digital preservation techniques. 24 individuals provided responses to each Likert scale question.

### Perceptions on the Amount of Labor Expended in Migration, Emulation, and Data Archeology

16 informants (67%) agreed or strongly agreed that both migration and data archeology are labor intensive. Informants, however, felt more strongly about the labor involved in data archeology than that of migration. Data archeology had a 12% higher response for strong agreement with the statement that data archeology is labor intensive than those who strongly agreed that migration is labor intensive.

A substantial percentage of informants neither agreed nor disagreed with the statement. Over half of individuals surveyed (54%) neither agreed nor disagreed that emulation is labor intensive, while 33% and 29% neither agreed nor disagreed that migration or data archeology, respectively, are labor intensive. Figure 9 illustrates the

exact counts of survey responses to the statement that migration, emulation, and data archeology (respectively) are labor intensive.

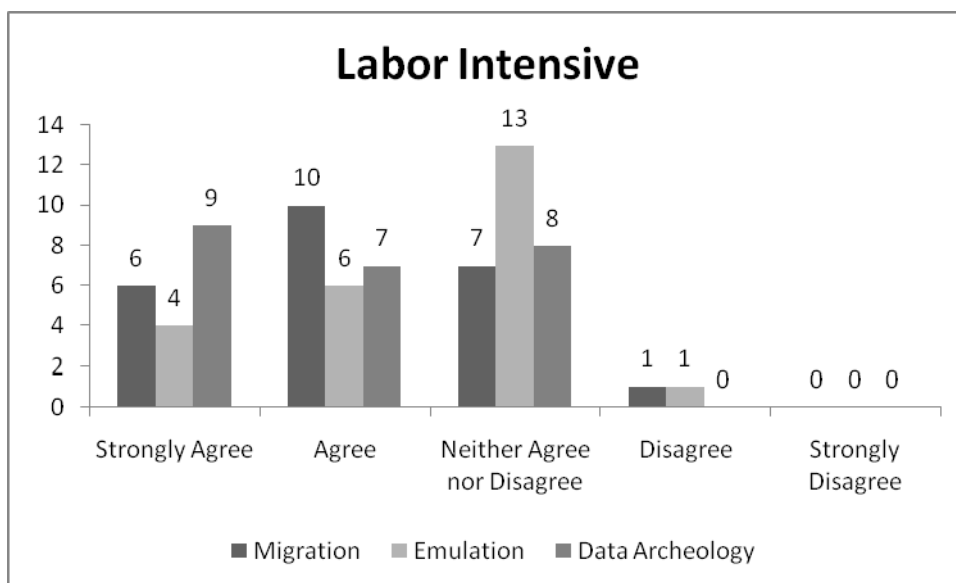


Figure 9: Institutional perceptions of the statement that migration, emulation, and data archeology (respectively) is labor intensive, by institution count.

#### Perceptions of the Financial Expense of Migration, Emulation, and Data Archeology

The survey results indicate that data archaeology is considered most expensive and emulation least expensive. Most respondents strongly agreed, agreed, or neither agreed nor disagreed with the statement of financial expense. Only one respondent disagreed that migration is financially expensive and no respondents disagreed that emulation and digital archeology are financially expensive.

67% of participants agreed or strongly agreed that data archeology is financially expensive. Respondents perceived migration to be more expensive than emulation. 54% agreed or strongly agreed that migration is expensive and 46% believed the same for emulation. Respondents seem to be more confident of the expense of data archeology than the costs of emulation or migration. 8 respondents strongly agree that data

archeology is financially expensive while 4 respondents strongly agree that migration and emulation are expensive. Figure 8 illustrates the exact counts of survey responses to the statement that migration, emulation, and data archeology (respectively) are financially expensive.

A substantial percentage of those surveyed neither agreed nor disagreed that migration, emulation, and data archeology are financially expensive. Over half of the responses neither agreed nor disagreed that emulation is financially expensive, 41 % felt the same way about migration, and 33% indicated the same response for data archeology.

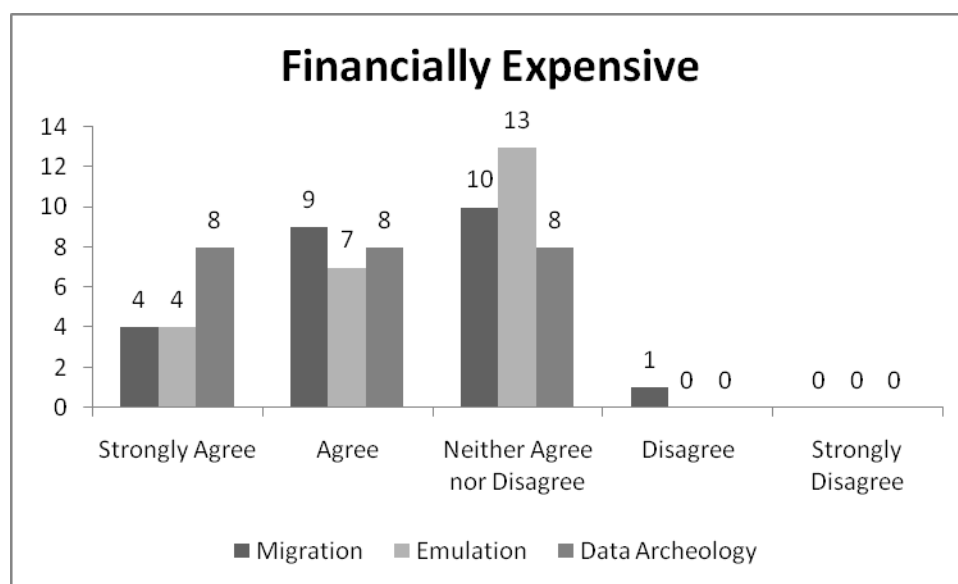


Figure 10: Institutional perceptions of the statement that migration, emulation, and data archeology (respectively) is financially expensive, by institution count.

### Perceptions of the Effectiveness of Migration, Emulation, and Data Archeology as a Digital Preservation Technique

There was very little distinction between the perceptions of the effectiveness of digital preservation techniques for text files versus those for image files. In fact, the

responses given to the statements regarding the effectiveness of migration, emulation, and data archeology for text files and the same for image files are either identical or only off by one as may be viewed in figure 11.

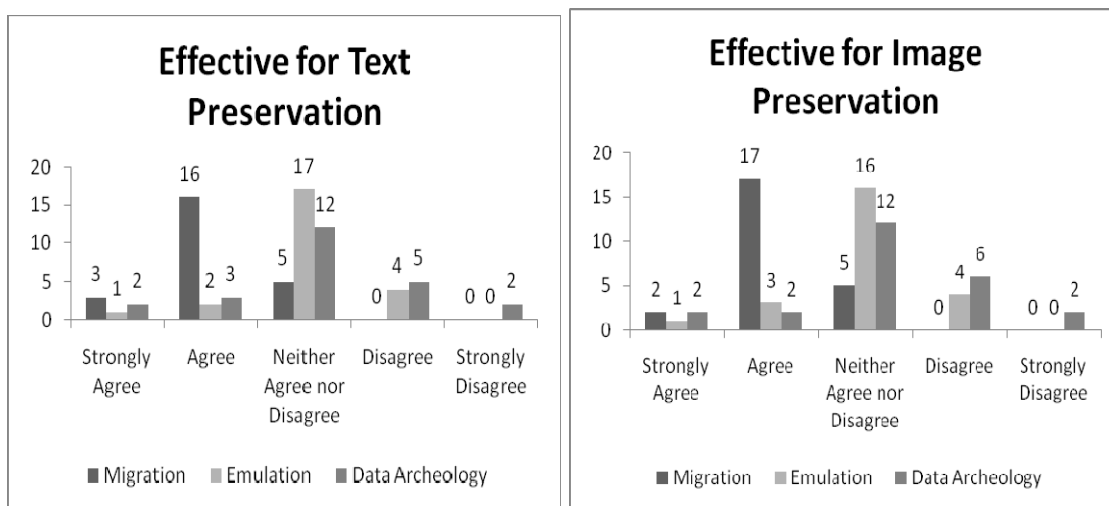


Figure 11: The nearly identical responses given to the statements that migration, emulation, and data archeology (respectively) is an effective preservation strategy for digital images and an effective preservation strategy for digital text.

Few respondents indicated that they felt strongly about the effectiveness of migration, emulation, or data archeology as a digital preservation technique and many of those surveyed neither agreed nor disagreed with this statement. The results suggest, however, that migration is perceived as a more effective preservation technique than either emulation or data archeology. 79% of respondents agreed or strongly agreed that migration is an effective preservation strategy for text and image files. A much lower percentage, however, either agreed or strongly agreed that emulation and data archeology are effective preservation strategies (12 -17% and 17-21% respectively, as the percentage varies based on type of file)

No respondent disagreed or strongly disagreed that migration is an effective digital preservation technique for image and text. However, 17% disagree that

emulation is an effective preservation strategy for digital text and images and 29 – 33% disagree that data archeology is an effective preservation strategy for digital text and images.

### Perceptions of the popularity of Migration, Emulation, and Digital Archeology

A high percentage of respondents neither agreed nor disagreed with the statements concerning the popularity of migration (62%), emulation (46%), and data archeology(67%). Most participants (50%), disagreed or strongly disagreed that emulation is widely implemented while 4% agreed or strongly agreed with this statement. The results suggest that migration is perceived to be most widely implemented. 29% of individuals surveyed agreed or strongly agreed that migration is widely implemented while 8% of individuals disagreed with this statement. No participant agreed that data archaeology is widely implemented. However, only 33 % disagreed or strongly disagreed with this statement. Figure 12 illustrates the exact counts of survey responses to the statement that migration, emulation, and data archeology (respectively) are widely implemented.

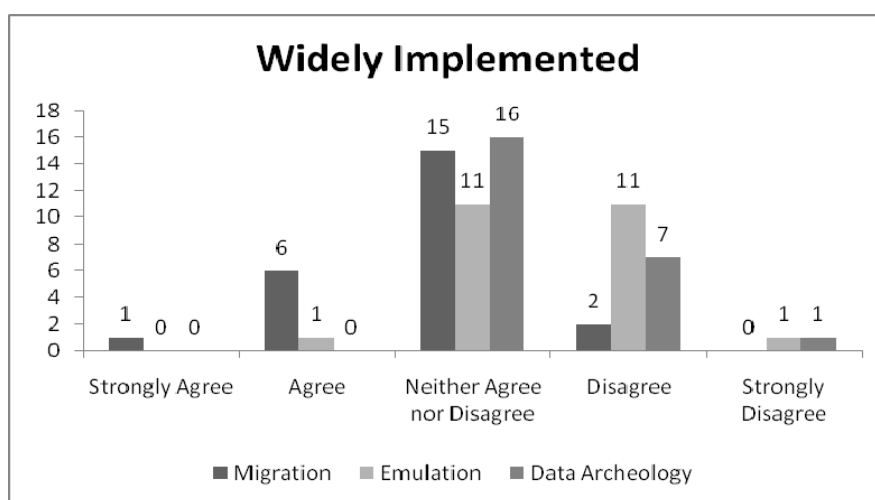


Figure 12: Institutional perceptions of the statement that migration, emulation, and data archeology (respectively) is widely implemented, by institution count.



### Neither Agree nor Disagree

With the exception of a substantial percentage of agreement with statements relating to migration (as discussed earlier), a large percentage of respondents neither agreed nor disagreed with each statement made about each digital preservation strategy. Figure 13 illustrates the number of ‘neither agree nor disagree’ responses given for all questions relating to the perceptions of migration, emulation, and data archeology.

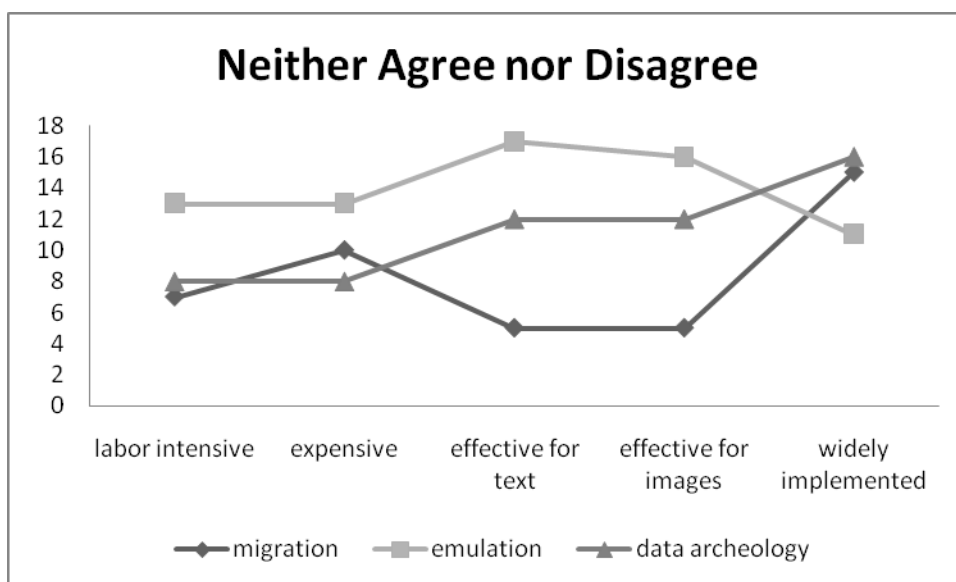


Figure 13: Counts of institutions that neither agree nor disagree with statements made about migration, emulation, and data archeology.

More respondents indicated that they neither agreed nor disagreed that emulation is an effective digital preservation strategy for images and text than those that indicated the same opinion for migration or data archeology. Emulation, overall, has the greatest amount of neither-agree-nor-disagree responses, except for perceptions of its popularity. Figure 14 reveals that emulation is perceived as the least widely implemented preservation technique of the three. These two results support the claim that ARL institutions are more familiar with migration and data archeology than

emulation. Migration has the least amount of respondents that neither agreed nor disagreed that it is an effective preservation strategy. Data archeology and migration have similar levels of respondents that neither agreed nor disagreed with the statements made about them.

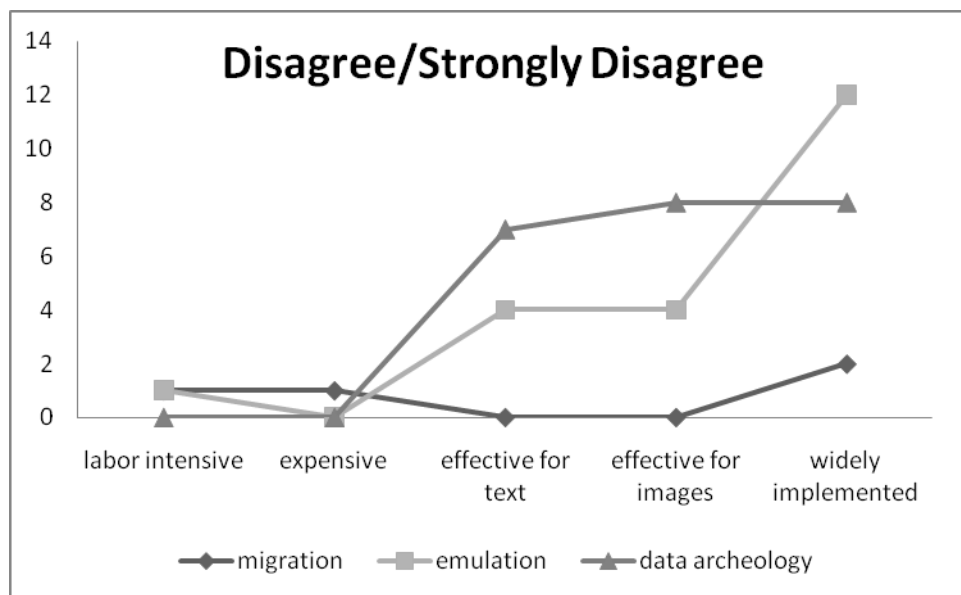


Figure 14: Counts of institutions that disagree or strongly disagree with statements made about migration, emulation, and data archeology.

### Participant Perceptions of Digital Preservation Techniques that their Institution has Adopted

A major research question for this study proposes to find if there is a relationship between the use of certain digital preservation techniques and the attitudes these institutions display regarding these particular techniques. The results indicate that there are less respondents that neither agree nor disagree with statements made about the digital preservation technique among the population that has engaged in this technique than among the general population. The degree to which this is true varies

based on the digital preservation technique in question. Each preservation technique is analyzed below.

The Perceptions of participants who have participated in migration regarding the strategy of migration

10 of 25 respondents indicated that their institution has engaged in migration of their digital image files, of these respondents 6 also have migrated their digital text files. The comparison of perceptions of institutions that have migrated and those that have not show the least amount of difference in the amount of the respondents in either group that neither agree nor disagree with statements made about migration. Respondents who indicate that their institutions has engaged in migration have 22% less neither agree nor disagree opinions regarding the financial expense of migration as oposed to respondents that have not engaged in migration. A large percentage of those who have participated in migration (80%) agree or strongly agree that it is financially expensive, compared to 55% of the general respondents who feel this way.

Figure 15 and 16 illustrate the percentage of survey responses (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree) to statements offered regarding migration for both the specaalized group that has engaged in this strategy, and the general group of respondents.

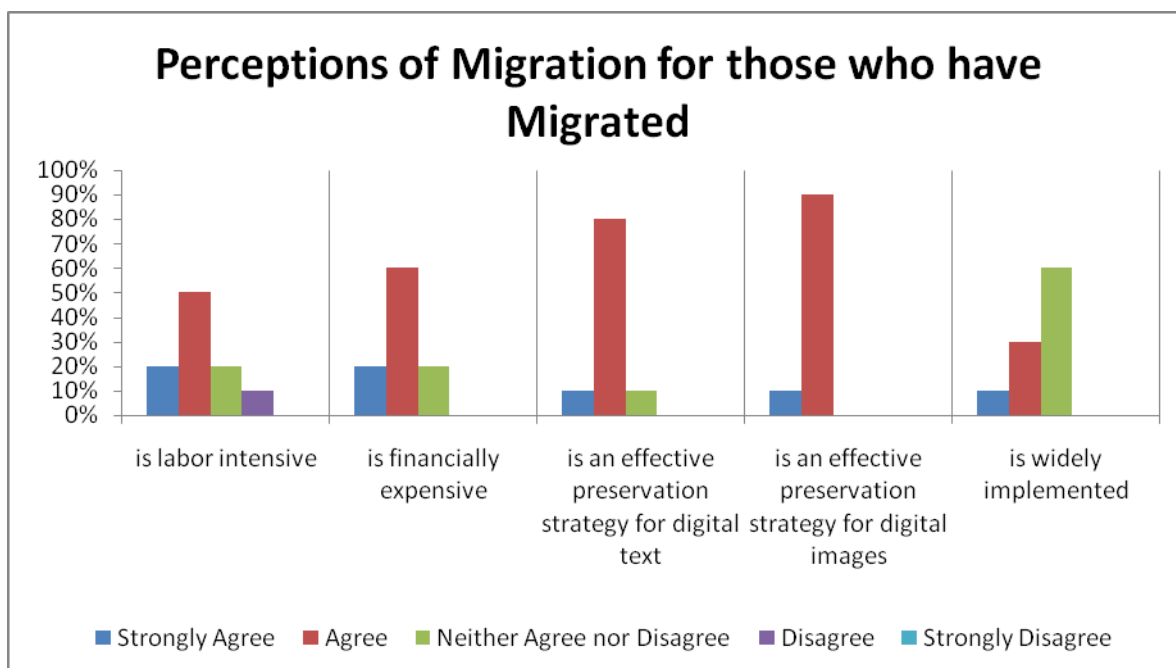


Figure 15: Perceptions of migration offered by respondents from institutions that have previously indicated that they have engaged in migration

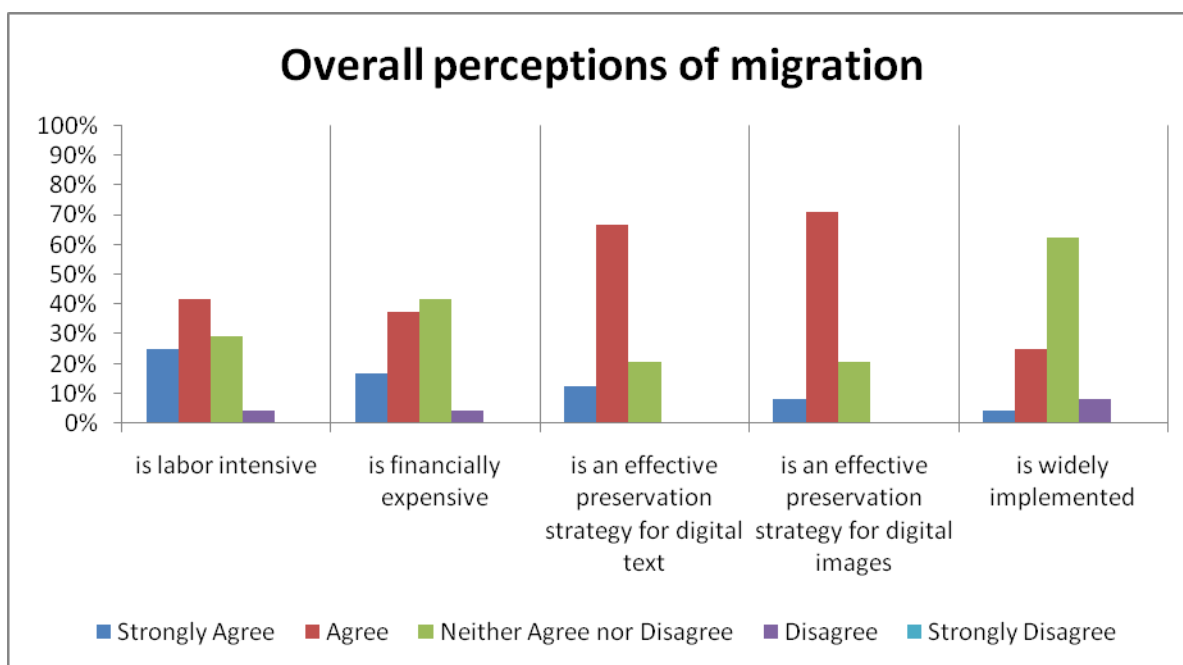


Figure 16: Overall perceptions of migration

The Perceptions of participants who have participated in emulation regarding the strategy of emulation.

Of 25 responses, two indicated that their institution has engaged in emulation of their digital image files or digital text files.

It is interesting to note that while over 50% of the the general population neither agreed nor disagreed with statements regarding the amount of labor, expense, and effectiveness of emulation, the population that has used emulation either agreed or disagreed with all of these statements. Both individuals agree that emulation is an effective preservation strategy for digital text and images. Both agree that emulation is financially expensive. One agrees that emulation is labor intensive, while the other disagrees. One agrees that emulation is widely implemented, while the other neither agrees nor disagrees.

Figure 17 and 18 illustrate the percentage of survey responses (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree) to statements offered regarding emulation for both the specailized group that has engaged in this strategy, and the general group of respondents.

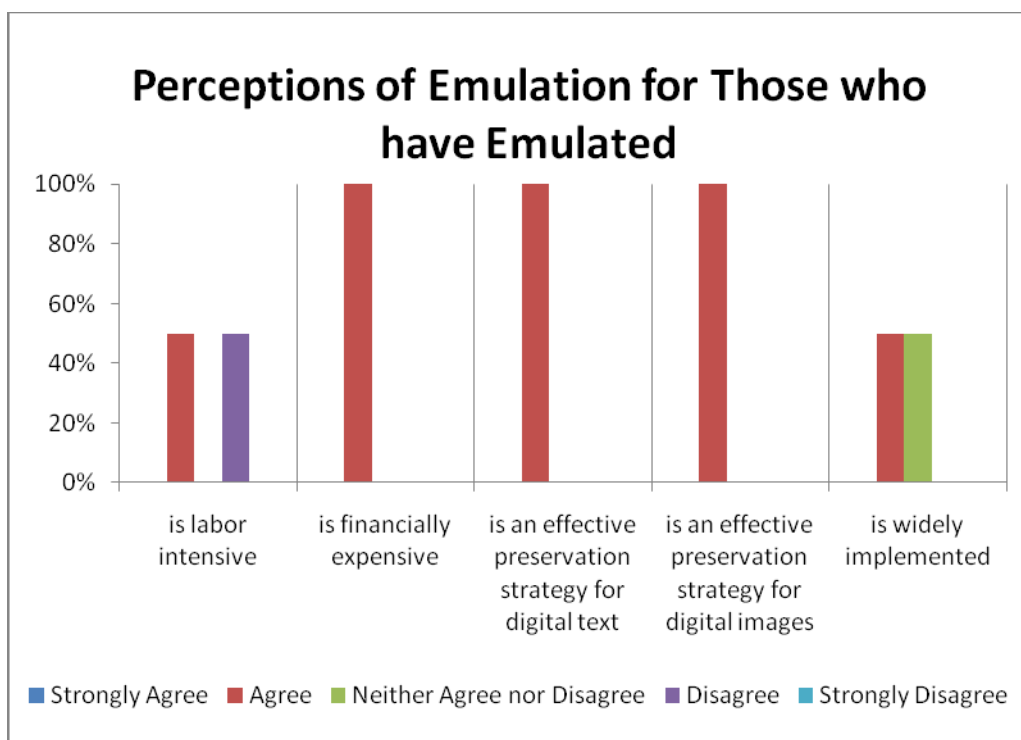


Figure 17: Perceptions of emulation offered by respondents from institutions that have previously indicated that they have engaged in emulation. N = 2.

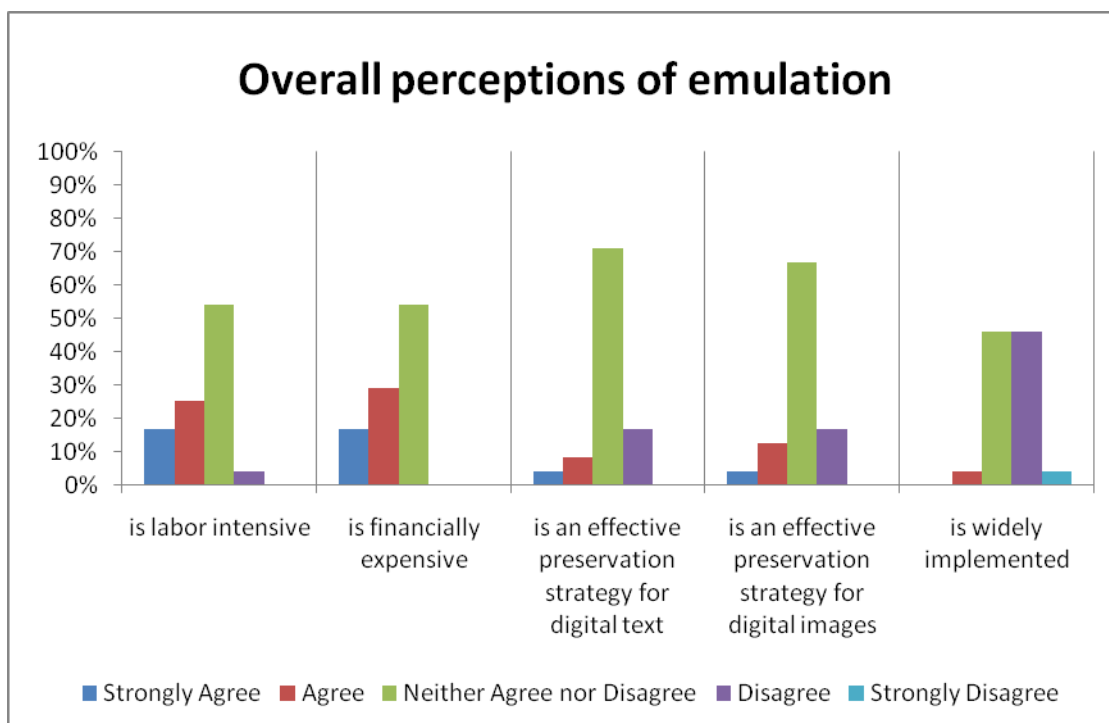


Figure 18: General perceptions of emulation.

The Perceptions of participants who have participated in data archeology regarding the strategy of data archeology.

Of 25 responses, 6 indicated that their institution has engaged in data archeology for digital images and 3 for digital text. Of the 6 individuals who had used data archeology on their digital image files, 3 had also used data archeology on their text files. Therefore group institutions who used data archeology on image files is inclusive of institutions who have used data archeology on text files.

Interestingly, in the population of individuals who had engaged in data archeology, there were no respondents that neither agreed nor disagreed with statements made about intensity of labor and financial expense of this preservation technique. 100% of individuals who indicated that their institution has used data archeology agree or strongly agree that data archeology is labor intensive and financially expensive. One-third of respondents in the general population indicated that they neither agreed nor disagreed with these statements, however no respondents in the general population disagree or strongly disagree that data archeology is either labor intensive or financially expensive.

Figure 19 and 20 illustrate the percentage of survey responses (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree) to statements offered regarding data archeology for both the specialized group that has engaged in this strategy, and the general group of respondents.

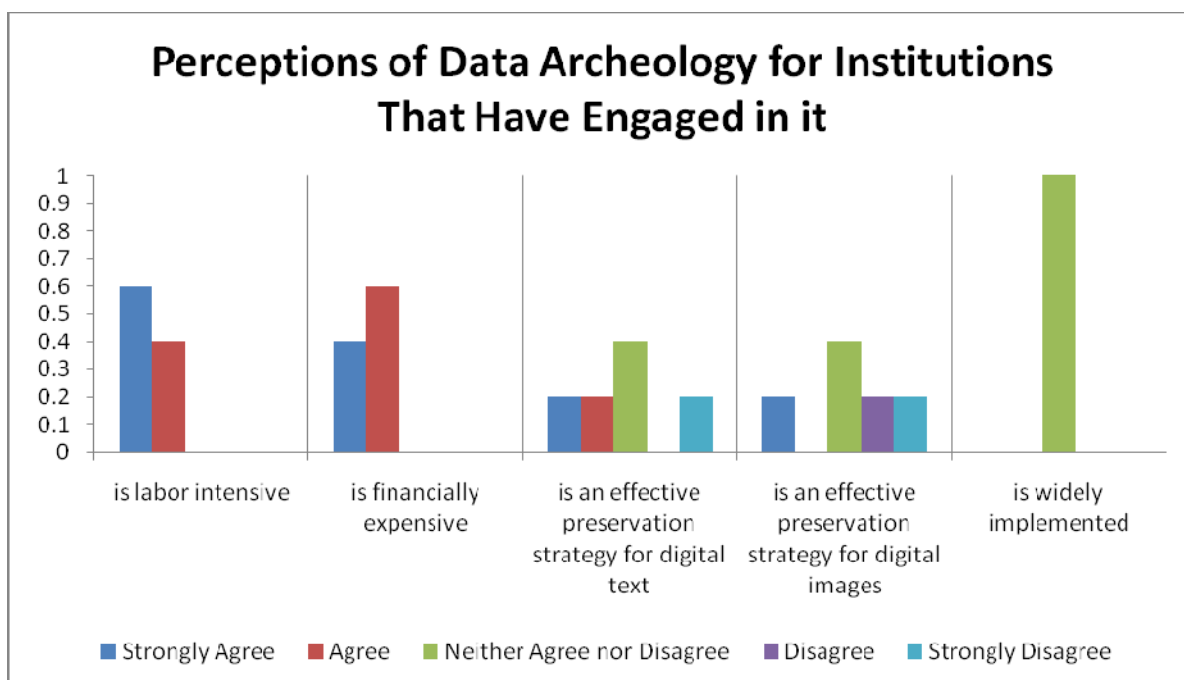


Figure 19: Perceptions of data archeology for institutions that have previously indicated that they have engaged in data archeology

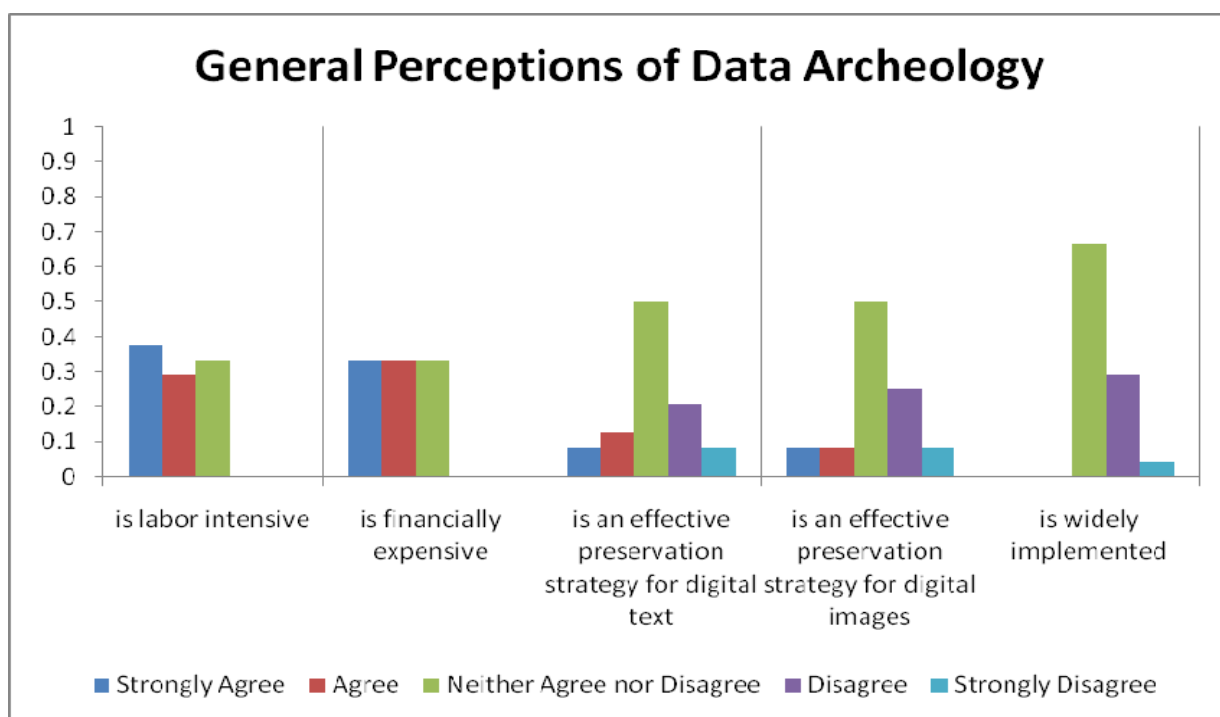


Figure 20: General perceptions of data archeology



## Discussion

The purpose of this research is to provide an exploratory study of the digital preservation activities and attitudes at ARL institutions with a special focus on the processes and perspectives of grant-receiving institutions.

The results suggest that both grant receiving and non-grant receiving institutions are more likely not to use migration, emulation, or data archeology than they are to use them. This result mirrors the surveys conducted by Simpson in the UK, and Pymm and Lloyd in Australia, both of which reported a low percentage of respondents involved in a digital preservation strategy). However, contrary to Pymm and Lloyd's 2005 findings in which no library surveyed was involved in a migration program (176), a large percentage of ARL institutions are engaged in migration.

The results suggest that there is no statistically significant relationship between the ability of an institution to receive a national digitization or digital preservation grant and the implementation of migration, emulation or digital archeology at that institution. Respondents from institutions that had received national digitization or digital preservation grants were just as likely to report that they engaged in migration, emulation, and digital archeology as institutions that had not received such grants. This seems inconsistent with the findings of the 2002 survey of US institutions and archival projects conducted by Cloonan and Sanett (which reports a high percentage of respondents that were engaged in a digital preservation strategy) as compared to the results of Simpson's survey of institutions in the UK and Pymm and Lloyd's Australian survey (which report a low percentage of respondents that were engaged in a digital preservation strategy). The institutions and projects that took the American survey all

employed a member of InterPARES and the disparity between the results of the American study and the UK and Australian study may be accounted for by the fact that the American study included only institutions and projects with a demonstrated commitment to digital curation. However, it appears that the ability of an institution to receive a national digitization or digital preservation grant does not provide the same litmus for the presence of digital preservation activities as participation in the InterPARES Preservation Task Force

ARL institutions hold much more than plain text and image files in their collections. Most also have audio and video files. Some even collect statistical datasets and websites. The surveys conducted in the UK and Australia suggest that the digital preservation practices for complex digital objects are similar to those for the simpler ones. However, my survey suggests that ARL institutions are more likely to practice a digital preservation strategy on digital image files than digital text files. Exactly why this may be true is unclear. One may speculate a variety of possible reasons: respondents may believe that digital image files are more vulnerable or valuable to the institution than text files, the literature may have been more successful in demonstrating the importance of digital preservation to image files, or the community may be more active in creating standardized file formats for images. While the survey results for digital text file formats were more varied, the results for digital image file formats were overwhelmingly consistent. Despite Buonora and Liberati's claim that JPEG 2000 has become the de facto standard for digital image archiving, it seems that TIFF remains the most common digital image file format in ARL institutions.

Nearly 2 in 5 institutions have migrated image files, 1 in 4 have used data archeology and 1 in 12 have emulated image files. In agreement with Kranch, Bullock, Granger and Rothenberg, migration is the most common digital preservation strategy used in ARL institutions. It is surprising that ARL institutions prefer data archeology (retro-archiving) to emulation due to the perceived high cost and amount of labor involved in digital archeology (Rothenberg). However, it seems that ARL institutions are well aware of the intensity of the labor and high financial cost involved with digital archeology. All of the ARL institutions, who had engaged in digital archeology previously, acknowledged that data-archiving is both labor intensive and financially expensive. Additionally, ARL institutions generally considered data archeology more expensive than migration or emulation.

The results suggest that the preservation perspectives of institutions that have engaged in various digital preservation techniques tend to differ from those which have not. Namely there are fewer respondents that neither agree nor disagree with statements made about specific preservation techniques that the institution has engaged in. The trend of neutrality in the perspectives of digital preservation techniques will be discussed later, but it may be important to prelude that discussion with the claim that has been illustrated by the preceding result – that the level of neutrality regarding statements made about various digital preservation techniques lessens as the participant has more personal experience with it.

The results display a high percentage of institutional neutrality over emulation, data archeology, and at times migration. This suggests that these techniques are not well understood. The literature provides many possible reasons why this is a valid claim.

Although there is much literature available, there is also much disagreement over which is the 'best strategy'.

It is important to emphasize the trend of neutrality regarding digital preservation perspectives as a large portion of institutions provided a neither agree nor disagree response to various statements made about digital preservation strategies. While the numbers suggest that ARL institutions consider migration more effective than emulation, there may be a large number of institutions who have not made the final determination, enough to sway the numbers either way. The survey offers insight into the perceptions of ARL institutions limited to one point in time. Among ARL institutions, data archeology was considered the most labor intensive and the most expensive. These results comply with Rothenberg's critique. Migration was a close second to data archeology in terms of labor and expense. However, migration was considered the most effective digital preservation strategy for digital image and text files. Emulation was marked by neutrality, except for perceptions of its popularity: it is considered the least widely implemented technique of the three.

It is also important to address the theme of neutrality in light of the individuals who were recruited to the survey. The head administrator of the library was asked to pass the recruitment message along to the individual that he or she considered most capable of providing the digital preservation practices of the institution. The job titles of these individuals support Hedstrom's statement that digital preservation has become the concern of more than just the archivist. The majority of job titles did not indicate specialization in archiving or preservation, but rather indicated specialization in digital technologies. One may speculate that the prevailing neutrality towards preservation

techniques may be the result of a lack of formal education in digital curation and archiving. However, the survey did not ask for educational background, and the data cannot test this hypothesis.

### **Closing Remarks**

Millions of dollars are being invested in digitization and digital acquisitions at libraries, but if these resources are not adequately preserved, they will become nearly irrevocably obsolete and thereby inaccessible. Certainly, we cannot expect grant funding agencies to fund repeat projects in the future to re-digitize or re-acquire what has been mishandled, in cases when that is even possible. Additionally, library users have grown accustomed to the convenience of digital materials, and one may expect that digital resources will become even more demanded by library users in the future. Adequately applying digital preservation principles to the vast quantities of digital materials in libraries will not only help ensure their long-term survival but it will also help libraries be more effective stewards of their resources. It is hoped that the findings of this research will inform academic librarians and digital archivists of the practices of academic libraries in the realm of digital preservation. Libraries may be, understandably overwhelmed with the issues of digital preservation. Many libraries may view digital preservation as beyond the scope of their resources. Therefore, a large priority of the research is to identify the attitudes that libraries have towards digital preservation techniques. Additionally, if digital preservation in archival institutions is truly as inactive as previous surveys indicate, perhaps knowledge of common perceptions will help experts frame their discussions in a manner that serves the academic library, and

remove the mystery associated with digital preservation in order to encourage the implementation of these techniques.

#### Suggestions for future research

While it may be clear that many institutions world-wide are not engaged in the digital preservation of their digital collections, the reasons why are less obvious. This study suggests that many individuals are not familiar with digital preservation techniques, but how does that play into library politics? McGovern reinforces the importance of organization in digital preservation through her metaphor of the three-legged stool. Are institutions not actively curating their digital collections because they do not know how? Is it because the administration is not convinced of the importance? Is it a matter of organizational support? Do they perceive their digital collections as still being so new that they do not yet have to worry about implementing digital preservation activities? Future research may be needed to reveal the reasons why institutions do not involve themselves in a digital preservation program.

## Works Cited

- Babbie, Earl. The Practice of Social Research. Belmont: Wadsworth, 2007.
- Balkestein, Marjan, and Heiko Tjalsma. "The ADA Approach: Retro-archiving Data in an Academic Environment." Archival Science 7.1 (2007): 89-105.
- Bhatnagar, Anjana. "Digitization in Academic Libraries." Information Studies 12.1 (2006): 35-54.
- Buckley, Robert. "JPEG 2000 – a Practical Digital Preservation Standard?" Digital Preservation Coalition February 2008: 26 March 2009.  
<<http://www.dpconline.org/docs/reports/dpctw08-01.pdf>>.
- Bullock, Alison. "Preservation of Digital Information: Issues and Current Status." National Library of Canada. 22 April 1999. 19 October 2008.< <http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes60.htm>>.
- Buonora, Paolo, and Franco Liberati. "A Format for Digital Preservation of Images: A Study on JPEG 2000 File Robustness." D-Lib Magazine. 14.7/8. 21 October 2008. <<http://www.dlib.org/dlib/july08/buonora/07buonora.html#1>>.
- Cloonan, Michele V., and Shelby Sanett. "Preservation Strategies for Electronic Records: Where We Are Now—Obliquity and Squint?" The American Archivist 65(2002): 70-106.
- Colin Webb, "Migration Trials: Migrating publications on floppy disk to CD-R." October 1997. 19 October 2008.  
<<http://www.nla.gov.au/nla/staffpaper/cwebb7.html>>.

- Granger, Stewart. "Emulation as a Digital Preservation Strategy." D-Lib Magazine. October 2000. 8 October 2008. <<http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/october00/granger/10granger.html>>.
- Hedstrom, Margaret, and Christopher Lee "Significant Properties of Digital Objects: Definitions, Applications, Implications." Proceedings of the DLM-Forum 2002, Barcelona, 6–8 May 2002. Luxembourg: Office for Official Publications of the European Communities, 2002. 218-227.
- Hedstrom, Margaret. "It's About Time: Research Challenges in Digital Archiving and Long-term Preservation." Washington, DC: The National Science Foundation and the Library of Congress, Feb 2003. 18 Oct 2008. <<http://www.si.umich.edu/digarch/NSF%200915031.pdf>>.
- Hedstrom, Margaret, Christopher Lee, Judith Olson, and Clifford Lampe. "The Old Version Flickers More: Digital Preservation from the User's Perspective." The American Archivist 69 (2006): 159-187.
- Kranch, Douglas A. "Beyond Migration: Preserving Electronic Documents with Digital Tablets." Information Technology and Libraries. 17.3 (1998). 20 Oct 2008. <[http://vnweb.hwwilsonweb.com/hww/results/results\\_single\\_fulltext.jhtml;hwwilsonid=FDXCTNJHGFNY5QA3DIKSFGOADUNGIIV0](http://vnweb.hwwilsonweb.com/hww/results/results_single_fulltext.jhtml;hwwilsonid=FDXCTNJHGFNY5QA3DIKSFGOADUNGIIV0)>.
- Mellor, Phil, Paul Wheatley and Derek Sergeant. "Migration on Request, a Practical Technique for Preservation." Research and Advanced Technology for Digital Libraries. (2002). 516-526. 8 October 2008. <<http://www.springerlink.com/content/752vmvw0g0w40dj2/fulltext.pdf>>



- Oltmans, Erik, and Nanda Kol. "A Comparison Between Migration and Emulation in Terms of Costs." RLG Digi-News 9.10 (2007). 8 Oct 2008.  
<<http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file959.html#article0>>.
- Peterson, Kit comp. "Standards Related to Digital Imaging of Pictorial Materials." Library of Congress. 2004. 10 Oct 2008.  
<<http://www.loc.gov/rr/print/tp/DigitizationStandardsPictorial.pdf> >.
- Pymm, Bob, and Annemaree Lloyd. "Dealing with Digital Collections: Interviews with the National Library and Selected State Libraries of Australia." Australian Academic & Research Libraries. 38.3 (2007): 167 – 179.
- Reference Model for an Open Archival Information System (OAIS) Blue Book. CCSDS 650.0.B-1. January 2002. 20 Oct 2008.  
<<http://public.ccsds.org/publications/archive/650x0b1.pdf> >.
- Rothenberg, Jeff. "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation." CLIR Reports. Jan 1998. 8 Oct 2008.  
<<http://www.clir.org/pubs/reports/rothenberg/contents.html>>.
- Stanescu, Andreas. "Assessing the Durability of Formats in a Digital Preservation Environment" OCLC Systems & Services. 21.1 (2005): 61-81.
- Simpson, Duncan. "Digital Preservation in the Regions Museums, Libraries and Archives Council." 2005. 20 Oct 2008.  
<[http://www.mla.gov.uk/resources/assets//M/mla\\_dpc\\_survey\\_pdf\\_6636.pdf](http://www.mla.gov.uk/resources/assets//M/mla_dpc_survey_pdf_6636.pdf)>.

## Appendices

### Appendix A: The Survey

1. Please identify the location of your institution (U.S. Canada Other)
2. Please indicate your job title.
3. Please provide the name of your institution. (Note: this information is used to identify duplicate responses and will not be included in any derived report).
4. Within the past 10 years, has the institution been involved in a national digitization and/or digital preservation grant?  
(Y N Applied but didn't receive funding Don't know)
5. Please describe the contents (e.g. text, images) and size of your institution's digital collections.
6. Please provide the following information for digital TEXT files:
  - a. What file format(s) do you use for the digital master copies of your digital text files?
  - b. On what physical media do you store your digital text collections?
  - c. Are you storing your digital text collections on physical media that are different from those to which you originally stored them?
  - d. What is the age range of these materials?
7. Please provide the following information for digital IMAGE files:
  - a. What file format(s) do you use for the digital master copies of your digital image files?
  - b. On what physical media do you store your digital image collections?
  - c. Are you storing your digital text collections on physical media that are different from those to which you originally stored them?
  - d. What is the age range of these materials?
8. Please indicate whether your institution has used migration in the management of your digital image files  
Yes No Don't Know
9. Please indicate whether your institution has used emulation in the management of your digital image files  
Yes No Don't Know
10. Please indicate whether your institution has used data archeology (extracting information from obsolete media – yourself or by hiring a vendor to do the extraction) in the management of your digital image files  
Yes No Don't Know

11. Please indicate whether your institution has used migration in the management of your digital text files

Yes    No    Don't Know

12. Please indicate whether your institution has used emulation in the management of your digital text files

Yes    No    Don't Know

13. Please indicate whether your institution has used data archeology(extracting information from obsolete media – yourself or by hiring a vendor to do the extraction) in the management of your digital text files

Yes    No    Don't Know

14. Please indicate your level of agreement with the following statements:

Migration:

is labor intensive

has a high financial cost

is an effective preservation strategy for digital text

is an effective preservation strategy for digital images

is widely implemented

Emulation:

is labor intensive

has a high financial cost

is an effective preservation strategy for digital text

is an effective preservation strategy for digital images

is widely implemented

Digital Archeology (extracting information from obsolete media – yourself or by hiring a vendor to do the extraction):

is labor intensive

has a high financial cost

is an effective preservation strategy for digital text

is an effective preservation strategy for digital images

is widely implemented