Beyond the Repository:

Integrating Local Preservation Systems with National Distribution Services

Authors:

Evviva Weinraub, Primary Investigator

Laura Alagna

Carolyn Caizzi

Brendan Quinn

Sibyl Schaefer

This project was made possible in part by the
Institute of Museum and Library Services
Grant LG-72-16-0135-16

Acknowledgments

The grant team would like to thank the many people who have contributed to this work: Gina Petersen, Michael Giarlo, Bertram Lyons, Mary Molinaro, Mike Ritter, Justin Simpson, David Wilcox, Andrew Woods, Declan Fleming, David Minor, Erin O'Meara, and all of those who took the survey and participated in interviews for this grant.

TABLE OF CONTENTS

Executive Summary ---- 4
Introduction ---- 5
Literature review ---- 6
Research Questions ---- 10
Methodology ---- 11
Data Files ---- 12
Survey and Interview Questions ---- 13
Findings ---- 13
- Background Information ---- 14
- Distributed Digital Preservation ---- 19
- Curating Distributed Materials ---- 24
- Versioning ---- 27
- Interoperability ---- 30
- Organizational Challenges ---- 31

Recommendations for Technical Solutions ---- 34
User Stories ---- 35
Conclusion ---- 36
Appendix A: Survey Questions ---- 38
Appendix B: Survey Visualization ---- 41
Appendix C: Interview Questions ---- 42
Appendix D: Interview Consent Form ---- 45
Bibliography ---- 46

Executive Summary

The “Beyond the Repository” planning grant investigated how local digital preservation practices and repository systems interoperate with distributed digital preservation (DDP) services. The grant team conducted a survey followed by in-depth interviews with selected survey respondents. The survey and interviews revealed both great diversity in how digital preservation is practiced and common challenges in the intersection of local repositories and DDP services.

The survey received 170 complete responses from a variety of organizations. Seventy-seven percent of respondents self-identified as academic institutions, but representatives from archives, government organizations, museums, non-profit organizations, and public libraries also responded to the survey. Survey respondents were nearly evenly split between those that identified as an administrator or department/unit head and those that identified as staff. The vast majority of survey respondents - 90% - reported that their institution had collected more than a terabyte of unique digital content, and 63% reported having collected fifty terabytes or less. Survey respondents reported using a variety of digital preservation and repository systems to manage their content; no one system was used by a clear majority of respondents.

The survey data reveals that most respondents (84%) are storing copies of their unique content in multiple locations. However, the number of copies stored varied among respondents: keeping two or three copies were the most common responses, but ten respondents reported keeping seven or more copies. In regards to where these copies are stored, survey respondents frequently indicated that their organizations pursue more than one storage strategy. Sixty-six percent keep copies in multiple locations onsite, but the cloud and DDP services are also common storage mechanisms. Of the survey respondents who use a DDP service, nearly half are members of the Digital Preservation Network, though several of these use DPN in conjunction with other services.

When asked about curation, almost half of the survey respondents indicated that they sent a subset of their data to a distributed repository (or offsite, or to the cloud). When these respondents were asked to rank the importance of criteria used to select the subset of materials sent off-site, the majority chose *Mandate* as the most important, followed closely by *Intrinsic value* and *Content type*. Sixty percent indicated that they have policies in place to guide selection of locally-held materials, but only 47% have similar policies for materials being sent to distributed systems. The interviews reflected this trend, with many interviewees commenting that they have criteria for selecting materials to be sent to offsite storage or DDP systems, but these are not necessarily articulated in policies.

Survey respondents and interviewees frequently cited lack of interoperability between tools and systems as a challenge. Many identified overspecialization of systems as a contributor to interoperability issues. Others described their systems as separate units with little integration between them, requiring manual processes and workarounds. One way this seems to

commonly manifest is the difficulty many respondents and interviewees had in tracking their content between systems.

This research uncovered a number of organizational challenges as well. A common theme in both survey responses and interviews was the lack of required funding or staffing for a robust digital preservation program. These factors were cited as the main reasons why respondents did not keep multiple copies of content in multiple locations, and as significant reasons why their organizations did not have digital preservation policies. Staff turnover was a challenge that was mentioned by many of the interviewees. Several mentioned struggles in retaining technical staff, and others noted that it was difficult to convince administrators to replace staff members who had left. In addition to the above challenges in integrating tools and systems, funding and staffing emerged as significant barriers to building robust digital preservation programs.

The grant team and the advisory board have coalesced around three recommendations after reflecting on the survey and interview results. The first recommendation is for the creation of a decision-making toolkit for choosing materials to send to DDP systems, which would help users with curation decisions and streamline digital preservation workflows. The second recommendation is to determine a shared BagIt profile for DDP systems, which would improve interoperability between systems. The third recommendation is a dashboard or similar tool that could be used to track content between systems. It is our hope that these recommendations are considered for any follow-on work from this project with the aim of improving DDP workflows and interoperability.

Introduction

Many cultural heritage institutions have been using a variety of solutions to preserve important digital content either generated or collected by their constituencies. Software repository systems such as Fedora and DSpace may provide some preservation functionality, but they pose risk factors such as lack of geographic diversity, lack of technological diversity, and loss of data related to human activities and systems failures. Distributed digital preservation systems, networks of nodes in geographically dispersed locations designed to perform preservation actions, aid in mitigating these risk factors. As institutions increase their local preservation capabilities, they often face challenges integrating local repository systems with distributed digital preservation systems. These challenges include: the management of multiple copies in multiple systems, selecting a subset of content from a corpus of digital materials that has already been deemed worthy of preserving; versioning digital objects and/or descriptive, administrative, technical, or preservation metadata; and the differences in data storage between systems.

Literature review

To more fully understand these issues, we have undertaken a review of the literature on Distributed Digital Preservation (DDP), as well as on the development of digital repository systems, related workflows and documentation.

Distributed Digital Preservation

While the first implementation of a distributed digital preservation was the Lots Of Copies Keep Stuff Safe (LOCKSS) initiative in the late 1990s, many authors trace the growth of later DDP initiatives and projects to the beginning of the National Digital Information Infrastructure and Preservation Program (NDIIPP), which first awarded grant funds for digital preservation projects in 2004 (Cruse and Sandore, 2009). Abby Smith's 2006 article on the NDIIPP examines the progress of the initiative over its first five years. According to Smith, the NDIIPP was charged with building “a national network of preservation partners who would collectively ensure long-term access to a rich body of digital content,” that would be supported by technical infrastructure constructed by the Library of Congress (Smith, 2006, p. 2). Though the NDIIPP is no longer an active program at the Library of Congress, it played an important role in funding early projects that developed preservation science or technical architecture, and projects that modeled or tested preservation strategies.

Several accounts of NDIIPP initiatives and projects, such as Chronopolis and the MetaArchive Cooperative, provide valuable information on the background of DDP. For example, in their 2009 article “The MetaArchive Cooperative: A Collaborative Approach to Distributed Digital Preservation,” Katherine Skinner and Martin Halbert describe the principles behind the founding of the MetaArchive Cooperative as well as how the organization functions. The authors argue that although the NDIIPP funded a wide variety of digital preservation initiatives, decentralized preservation networks like MetaArchive “have the best chance of accomplishing their aims in an affordable and sustainable manner through collaborative efforts” (Skinner and Halbert, 2009, p. 383). Building on this article, Skinner, along with co-author Matt Schultz, published "A Guide to Distributed Digital Preservation" in 2010. The most comprehensive source on DDP available, the document is organized into chapters intended to work on their own or together. It examines many aspects of DDP in great detail, including its theoretical basis, technical and organizational considerations, content issues (such as selection, preparation, management, ingest, monitoring, and recovery), cache and network administration, and copyright issues.

The development of distributed digital preservation systems is well documented. Private LOCKSS Networks (PLNs) such as the MetaArchive Cooperative and the Council of Prairie and Pacific University Libraries (COPPUL) have grown and matured for more than a decade. Articles and documentation on the development, policies, and technological basis of these networks are readily available. For example, representatives from MetaArchive, COPPUL, and the Alabama Digital Preservation Network came together in 2009 to produce “Distributed Digital Preservation: Technical, Sustainability, and Organizational Developments” (Walters et al., 2009). The authors describe how each network is organizationally structured, how each has

achieved sustainable growth, and the technical achievements of each. The article suggests that each distributed digital preservation network has taken a slightly different approach in these areas, but they “share the same mission of building successful means to preserve digital assets of scholarly, research, and cultural value” (Walters et al., 2009, p. 205).

Similarly, "Chronopolis Digital Preservation Network" (Minor et al., 2010) is a report on the first year of the Chronopolis network that examines both the theory behind the network's development, as well as its core technologies and services. The report concludes with future developments under investigation for Chronopolis, intended to strengthen ties between the network and the curation processes of member organizations. There are also many resources available on DDP systems that have emerged more recently, such as the Academic Preservation Trust and the Digital Preservation Network. For example, Martha Sites' 2013 article "The APTrust Story: a collaborative model for digital preservation" describes the problems APTrust is intended to address, as well as its founding, organizational and technical structures, and its role as a replicating node of DPN. Sites emphasizes the collaborative nature of APTrust and DPN, concluding that their “hope is to leverage the benefit of doing [digital preservation] together” (Sites, 2013, p. 14).

Finally, Nathan Hall and Michael Boock recently undertook an analysis of nine different DDP service providers. In their 2017 article “Environmental Scan of Distributed Digital Preservation Services: A Collective Case Study,” Hall and Boock reviewed the operations and service models of MetaArchive, APTrust, DPN, TDL, DuraCloud, Preservica, Chronopolis, Rosetta, and Arkivum. They evaluated each based on a variety of criteria and provide detailed comparisons of the nine systems in terms of their organizational and technical aspects. As such, Hall and Boock’s work is the most comprehensive comparison of these DDP service providers to date.

Local repository development

The literature on local repositories is similarly robust. Documentation on repository systems is often made available by service providers (see, for example, the documentation for Fedora or DSpace made available on the DuraSpace wiki, or the 1998 paper from Sandra Payette and Carl Lagoze describing the Fedora architecture). Academic articles and other resources on various aspects of digital repositories are also plentiful. In 2006 Richard Jones wrote about the development of digital repositories, tracing the emergence of institutional repositories to disciplinary repositories and the archives from which they grew. In 2009 the literature on this topic was substantial enough for Charles W. Bailey, Jr. to publish his “Institutional Repository Bibliography,” listing sources on digital preservation, metadata, software, and other issues related to institutional repositories, which he updated through 2011 (Bailey, 2011). Other authors specifically addressed the issue of digital preservation in the context of digital repositories: in 2002 William G. LeFurgy outlined a model for levels of digital preservation services for digital repositories. Greg Tananbaum's 2009 keynote address on the past, present, and future of institutional repositories at the Association for Library Collections & Technical

Services (ALCTS) Midwinter Symposium prompted the ALCTS to conduct a series of webinars on repository-related topics between 2010 and 2012.

More recently, efforts to address changes in the local repository landscape have emerged. As part of its Spring 2017 meeting, the Coalition for Networked Information (CNI) held two Executive Roundtable sessions on the topic of “Rethinking Institutional Repository Strategies.” In his summary of the roundtable discussions, CNI Executive Director Clifford Lynch acknowledged that much has changed since institutions began developing their institutional repositories: disciplinary repositories have become increasingly sophisticated, the rise of “public access” funder mandates and the “open access” movement have led to confusion over the role of institutional repositories, and repository platforms and the overall systems landscape have evolved substantially (Lynch 2017). CNI published a report expanding on Lynch’s talk that provides much more detail on the roundtable participants’ observations. Many of these touch on the lack of consensus on what a repository should be and what it should do. For example, organizations “are still debating whether a repository should be focused on discovery, access, and/or preservation” (CNI, 2017, p. 7). The report concludes by noting that participants urged CNI leadership to work towards a “global vision for repositories” that would lead to improved linkages between repositories and ultimately a global network (CNI, 2017, p. 12).

Interoperability efforts

Despite the wealth of literature on both distributed digital preservation systems and digital repositories, a gap exists in the connections between the two. Few sources mention links between local repositories and distributed systems, and many of the sources that exist on this topic are project reports on current work rather than research and analysis. However, there are several notable exceptions to this in the literature.

First is the expansion of the Open Archival Information System (OAIS) model into the Outer OAIS-Inner OAIS (OO-IO) Model, which takes multiple systems into account. In their 2014 article describing this conceptual framework, Eld Zierau and Nancy Y. McGovern argue that the OO-IO Model is necessary for DDP because the OAIS Reference Model only briefly discusses interoperability between two OAIS systems. The OO-IO Model “supports audit requirements within distributed digital preservation environments by elaborating the relationships and roles of functional entities and their functions within and between relevant OAIS’s” (Zierau and McGovern, 2014, p. 3). In the OO-IO Model, DDP services may serve as an Inner OAIS system. In this way, the OO-IO Model simplifies organizational and conceptual challenges of DDP that involves several organizations.

In addition to the conceptual framework for interoperability established by the OO-IO Model, there have been several initiatives working towards technical interoperability in digital preservation practice. First is the significant work by Priscilla Caplan, William Kehoe, and Joseph Pawletko on the "Towards Interoperable Preservation Repositories" (TIPR) project between 2008 and 2010. Caplan, Kehoe, and Pawletko examined the issues involved in transferring complex digital objects to distributed preservation systems, and developed a

Repository eXchange Package (RXP) that would facilitate transfer between dissimilar preservation repositories. The RXP is composed of four required XML files and the content that they describe. The XML documents include two METS descriptors that outline the sender, rights, and representations of the RXP and the representation of the sender's Dissemination Information Package (DIP); as well as two PREMIS documents that contain digital provenance information.[1] The RXP may also include optional files such as a signature file containing the sender's private key. The RXP specification, along with reports summarizing this IMLS-funded project, is available via the Florida Center for Library Automation (http://wiki.fcla.edu:8000/TIPR/). The project team has also published several articles describing the RXP and their work (for example, see Caplan et al., 2010).

Another important project addressing interoperability issues is the ArchivesSpace-Archivematica-DSpace Workflow Integration Project at the University of Michigan's Bentley Historical Library. Funded by the Andrew W. Mellon foundation, staff at the Bentley created an end-to-end workflow for digital archives. The project is described in detail in a 2017 article by Max Eckard, Dallas Pillen, and Mike Shallcross. While the work done for this project is a significant step towards interoperability in the digital preservation field, it does not address issues presented by DDP systems. The workflow created at the Bentley ends at the point of ingest into Deep Blue, the University of Michigan's instance of DSpace. However, this project could offer methods for approaching problems of interoperability between local repositories and distributed preservation systems. In addition to streamlining the digital archives workflow, the basic goals of the project were to "[f]acilitate the creation/reuse of...metadata across preservation and management systems," and to find solutions that would not be strictly for the benefit of the Bentley; solutions that are "flexible and scalable...modular, so that other institutions may adopt some, none, or all of the development features; and based upon open standards so that other tools and/or repository platforms could be integrated" (Eckard et al., 2017, p. 2). The goals of this integration project could easily be used as a model for a similar project promoting interoperability between local and distributed preservation.

Caplan et al. and Eckard et al. have done significant work in examining how systems and repositories could be made more interoperable, yet a dearth of literature about integrating local repositories with distributed preservation systems remains. Workflow documentation for preserving born-digital collections is one area that could be expected to document interoperability, and a few do. For example, of the thirteen institutions that have made their workflows available on the BitCurator Consortium website, five of them mention preservation storage, whether local or distributed (BitCurator Consortium, 2017). These five mention storage only generally; the workflows include steps like "move digital objects to repository." Only the workflow from Purdue University contains steps for preparing materials to be sent to a DDP service: according to the document, staff at Purdue use Bagger to package content for ingest into MetaArchive (Purdue University Workflow, 2016).

[1] The RXP includes metadata documents expressed using the Metadata Encoding & Transmission Standard (METS) and PREservation Metadata: Implementation Strategies (PREMIS) standards. For more information on these schema, see https://www.loc.gov/standards/

Clearly, more work needs to be done in examining interoperability between local repositories and distributed digital preservation systems, both in terms of researching how organizations are currently dealing with these challenges, and how interoperability could be improved to facilitate long-term digital preservation. The “Beyond the Repository” grant team investigated these issues through the lens of several research questions.

Research Questions

As evidenced by the literature review, local repository systems and distributed preservation systems have developed independently of one another, and both the literature and actual systems contain few overlaps. Though distributed digital preservation is necessary to mitigate risk factors such as lack of geographic or technological diversity, it can be costly and often a subset of an institution’s entire repository corpus must be selected. For example, Digital Preservation Network members can deposit up to five terabytes of data annually before incurring additional per-terabyte charges beyond their initial membership fee. This selection must be made prior to ingest into the distributed system.
How does an institution further select materials out of a collection of materials which have already been deemed as valuable? How does one curate objects to ingest into a long-term dark preservation system?

By their nature, dark distributed systems are not active; once data are deposited, updates and deletions should be infrequent. This can be problematic, given the tendency for digital collections, and especially the metadata describing them, to evolve over time. Part of the problem arises from the way digital collections or assets are acquired, as collections are often accumulated over time rather than fixed entities. Descriptive metadata changes happen at various points in time, and this dynamic content needs to fit within a logical and technical framework designed for managing static content. Versioning functionality may exist in local repository systems but does not necessarily translate to distributed systems. The management of multiple copies in multiple systems is problematic in numerous ways: systems use different identifiers and local versions are more dynamic and mutable than distributed ones.
How does versioning of objects and metadata play out in a long-term dark preservation systems and can these actions be automated?

Lastly, in addition to difficulties associated with managing multiple copies in multiple systems, the actual storage of data in systems differs. For example, Fedora 4 stores its binaries in a pair-tree structure on a file system and stores RDF metadata in an underlying database, but distributed systems like Chronopolis and DPN repackage stored data as files in the BagIt File Packaging Format (Kunze, et al., 2016).
What implications do these differences have for the restoration of data into the original system? How can systems that store data differently be made more interoperable?

Methodology

Data for this planning grant was gathered through two methods: a widely distributed survey and in-depth interviews of twelve selected survey respondents. Throughout the research process, the grant Advisory Board -- made up of representatives from Artefactual, AVPreserve, Chronopolis, the Digital Preservation Network, Fedora/DuraSpace, and Samvera -- offered guidance and feedback. The project team and Advisory Board met in person twice during the research process and communicated via conference calls and email.

Survey questions were developed by the grant team over a series of meetings and were then vetted by the Advisory Board. The grant team created the survey, disseminated it, and collected the resulting data using the Qualtrics data collection management system. The survey was distributed in two phases: it was sent via Qualtrics directly to targeted recipients, and then one week later links to the survey in Qualtrics were posted widely on 18 different professional listservs and Google groups.[2] The initial focused distribution was designed to overcome an inherent selection bias in the research design. The grant team recognized that it would be impossible to reach every organization doing digital preservation, and that using professional association mailing lists to disseminate the survey would likely result in a respondent group in which large academic research libraries were disproportionately represented. By directly inviting individuals who represent a variety of types of libraries, museums, archives, and preservation solutions to complete the survey, the grant team sought to ensure representation from a variety of institution types. Therefore, the survey was initially sent directly to 43 individuals and five DDP networks suggested by the Advisory Board or selected by the grant team because their responses would help diversify the overall data set. Reminder messages were sent the following week, and the survey was closed just under four weeks from when it was initially sent to targeted recipients.

After the survey was closed, the data was analyzed. Two members of the grant team reviewed the survey questions that included a free text box where respondents could write in an answer. Common themes among free text answers were highlighted and the number of answers that corresponded to each theme were tallied (specific themes the team drew from the free text answers will be discussed below in Findings).

Next, the grant team selected interviewees from the survey respondents. In order to be considered for a follow-up interview, respondents had to indicate a willingness to discuss their institution's digital preservation practices further and indicate that their organization currently keeps multiple copies of content in multiple places. The research questions guiding this work

[2] These include listservs for the American Library Association-Digital Preservation, the American Library Association Library and Information Technology Association, the Association of Southeastern Research Libraries, the BitCurator Consortium, the Coalition for Networked Information, Code4Lib, dh+lib, the Digital Library Federation, the Digital Preservation Coalition, Jisc Digital Preservation, the Library of Congress Digital Preservation Outreach and Education network, the National Digital Stewardship Alliance, the Open Preservation Foundation, the PREMIS Implementers Group forum, the Preservation and Archiving Special Interest Group, the Society of American Archivists, and Web4Lib. Google groups include the Digital Curation Google group and the BitCurator Users Google group.

focus on long-term dark storage systems and how different systems interoperate; interviewing respondents who use long-term dark storage systems or who store data in multiple systems would be essential to answering these questions. From respondents who met these criteria, the grant team selected fifteen[3] individuals whose survey responses indicated that they would provide useful insights for this research. The grant team also intentionally selected interviewees representing a diverse group of organizations, including a public library, a museum, and a state library, as well as public and private universities.

Interviewees were offered the opportunity to be interviewed in person at the Open Repositories 2017 Conference, or virtually via the BlueJeans video conferencing software. Prior to each interview, the participants gave their consent to be interviewed by submitting an online consent form via Qualtrics (see Appendix D). All of the interviews were recorded, and the audio recordings were transcribed using services provided from Rev.com. The grant team analyzed the interviews by coding the transcripts; highlighting areas where interviewees had mentioned certain themes allowed the team to compare discussions of each topic across the interviews. Some themes were common across all interviews because of the interview questions, such as curation, versioning, workarounds, tracking content, and interoperability. The grant team analyzed other themes that emerged from the interviews as well: many interviewees mentioned packaging content, staffing issues (such as lack of staff and turnover), and their relationship with information technology departments.

Data Files

The grant team used Qualtrics and an exported CSV spreadsheet to analyze the survey data for this report. To allow data reuse and long-term accessibility, the survey response data will be deposited into Northwestern University's institutional repository. Because it is impossible to fully anonymize the interviews, the transcripts and coded spreadsheet will not be included in the deposit.

The grant team decided not to include any partial responses to the survey on the grounds that we should respect the wishes of those who chose not to complete the survey. Furthermore, removing these partial responses would eliminate the possibility of including duplicate data, since it was impossible to know if a respondent had returned later to finish the survey. The 170 responses analyzed in this report include only those who reached the end of the survey, though they may have chosen not to answer every question.

To preserve survey respondents' anonymity, several fields of data automatically collected by Qualtrics were deleted. These include the respondents' IP addresses and approximate latitude and longitude. Responses including contact information were also deleted and survey responses were reviewed to ensure that identifiable information was removed from free text fields.

[3] Of this group, the grant team completed interviews with only twelve. One respondent chose not to be interviewed, and two did not respond to interview requests.

Survey and Interview Questions

The survey questions were designed to address the above research questions and to yield additional information that could be analyzed to give further insights on obstacles to building robust digital preservation programs. The survey posed a series of questions to gather information on the respondent's current digital preservation environment, including how much content had been collected and how much the respondent expected their organization to collect within the next year. A number of survey questions were dependent on how the respondent answered previous questions; for example, if a respondent answered that their institution did not keep multiple copies of content in multiple locations, then they would not see successive questions related to distributed digital preservation practices (questions dependent on skip logic are indicated in Appendix A and visualized in Appendix B). The survey contained a number of questions on distributed digital preservation, including questions on selection criteria for materials being sent to distributed storage, as well as versioning practices and preservation policies for locally-stored and distributed content. Respondents were also asked to describe what is missing in the technology they are currently using in order to better understand common challenges with existing systems. Finally, the survey posed demographic questions such as location of the institution and the role of the respondent within the organization. Respondents were asked to provide their name and contact information if they were interested in participating in a follow-up interview.

The interviews were structured to dig deeper into each interviewee's survey responses and provide a more detailed picture of the digital preservation environment, policies, and workflows at each organization. Interview questions were customized for each interviewee depending on his or her survey responses; not all interview questions were applicable to every interviewee (all possible interview questions can be found in Appendix C). The interviews started by asking the interviewees to describe the digital preservation program at their institution, including the number of staff involved, their role in the program, and what systems or services are used for digital preservation. The interviewees were then asked to describe their storage environment and how their local systems interoperate with their DDP system, if applicable. Additional questions about versioning practices, curation, and preservation policies were posed in the interviews so interviewees could elaborate more on these topics than in their survey responses. Finally, the interviews ended with the interviewers asking for any additional thoughts or comments, so that interviewees had a chance to express any details that they thought were relevant to the grant research that had not yet been discussed.

Findings

This section of the report is organized by themes from the initial research questions with the addition of "organizational challenges," which emerged as a new theme from the data collected. Discussion of findings on each theme begins with the text of original survey questions relevant

to that theme. Findings from both survey data and the interviews are analyzed for the following major themes:

- Background Information
- Distributed Digital Preservation
- Curation
- Versioning
- Interoperability
- Organizational Challenges

Note: Not all survey respondents answered all survey questions. Throughout this report, proportions are calculated as a percentage of those responding to the specified question.

Background Information

Digital Preservation Program

After an introductory section (Q1), the survey began with background questions to provide general information on the digital preservation programs at respondent organizations. All survey respondents were shown these questions, though not all of the 170 respondents to the survey chose to answer each question.

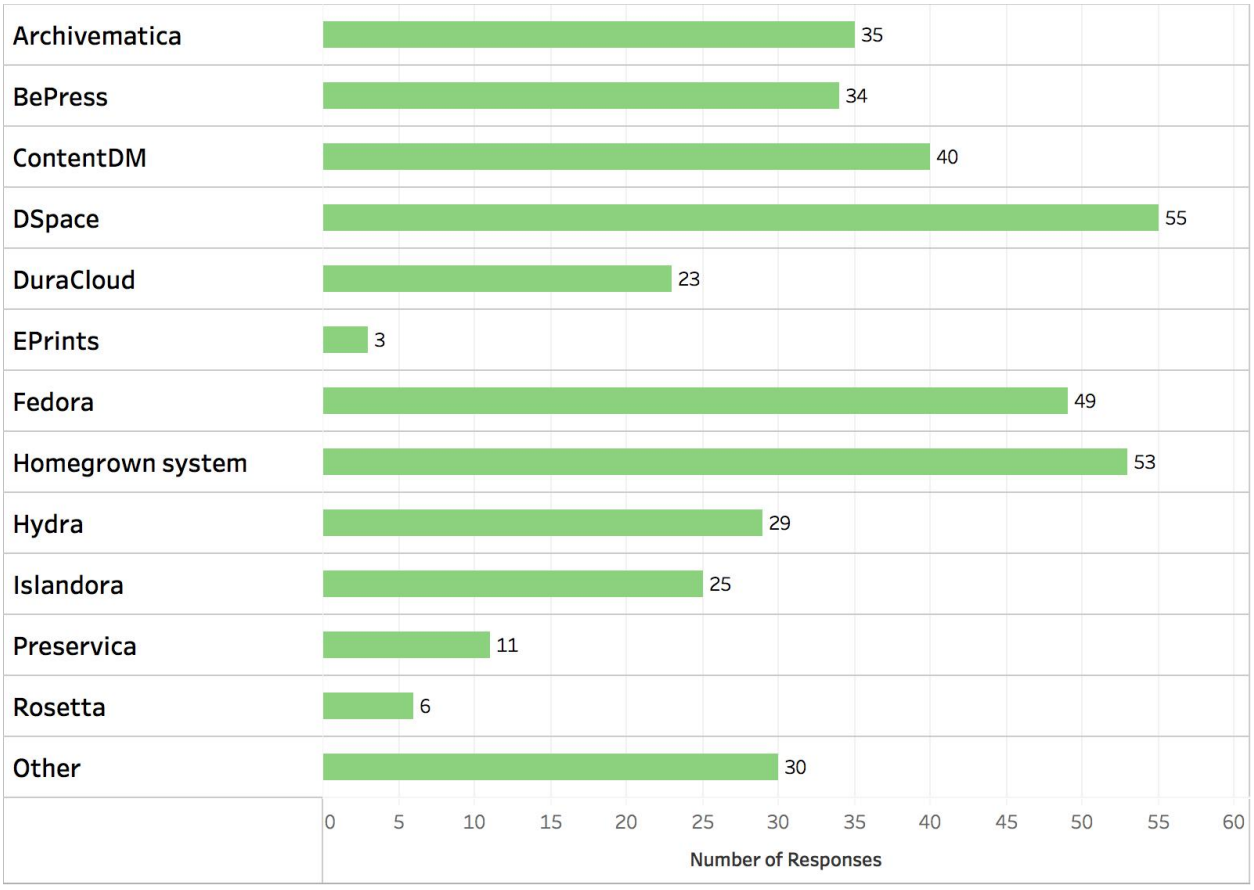
- Q2: Which of the following digital repository or digital preservation systems are used at your institution? Select all that apply.
- Q3: Approximately, how many terabytes of unique content has your institution collected?
- Q4: In the next year, how much *additional* data do you expect your institution to collect?

Survey respondents reported collecting varying numbers of unique digital content. Out of 156 respondents who answered Q3 concerning the amount of content they had collected, 142 (more than 90%) had collected more than a terabyte. The majority of respondents (99, or 63%) reported having between one and fifty terabytes. However, a significant proportion of respondents (nearly 20%, or 29 of those who had answered the question) had more than 100 terabytes of unique content. Responses to Q4 in the survey indicate that respondents expect to see their unique content grow quickly in the near future. When asked to estimate the growth rate of their data over the next year, many of the 156 who responded to Q4 indicated that they expect to see their content double or more. Of the 78 who reported that they currently hold between one and ten terabytes of unique content, 40 said that they expect to acquire between eleven and fifty terabytes in the next year; 12 of the 30 who currently hold between eleven and fifty terabytes expect to acquire more than 100 terabytes in the next year.

To manage this quickly growing content, survey respondents to Q2 reported using a variety of digital preservation and repository systems (see Figure 1). No single system held a clear majority, and between the options listed on the survey and the “other” category, respondents cited using 23 different systems. Of the thirty respondents who chose “other,” 26 explained their choice using the free text box for this answer. These answers indicate that organizations are

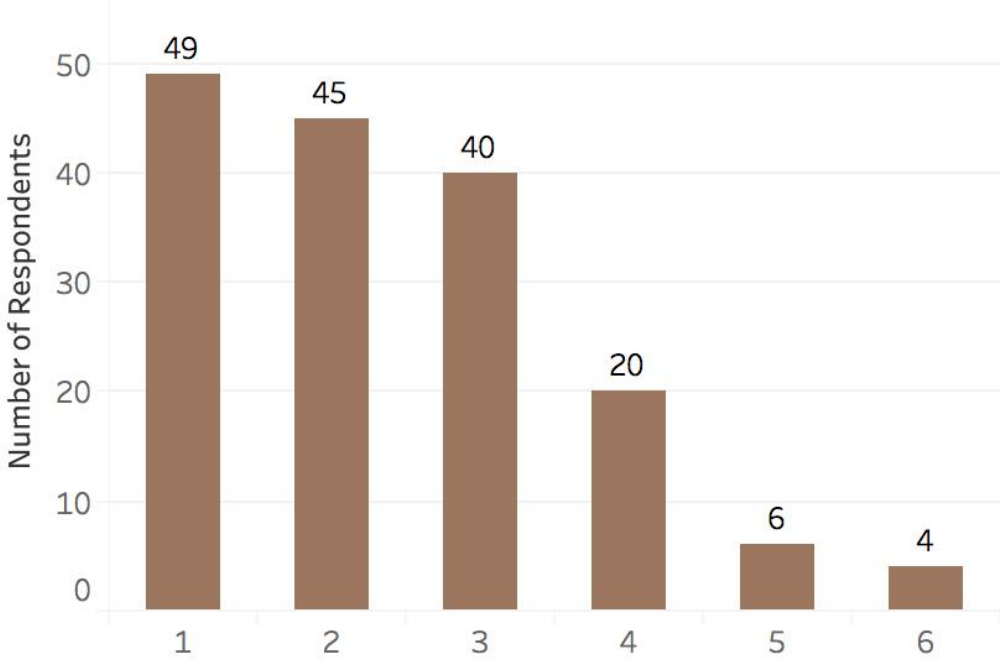
using a broad range of tools for digital preservation, and that many have programs that are in a state of transition. Both LUNA and Omeka were chosen each by three different respondents, and DigiTool from ExLibris was chosen by two; nearly a third of respondents who chose “other” for this question are using tools for digital preservation that have some digital preservation functions but are not designed to be digital preservation systems. In addition to this, four respondents who chose “other” described their digital preservation program in aspirational terms: “we are hopefully switching to a blend of Fedora/Hydra and Archivematica on the back end this year,” and “moving to Preservica” are typical of these answers.

Figure 1. “Which of the following digital repository or digital preservation systems are used at your institution? Select all that apply” (164 respondents)



The survey data about the digital preservation and repository systems in use reveal several trends. First, it’s clear that many survey respondents are using multiple systems. When asked which systems are used at their institution, respondents were allowed to select all options that applied. As a result, there are 393 responses to this question for 164 respondents who answered the question. This indicates that many respondents chose at least two systems; in fact, forty respondents chose four different systems, and some chose as many as five or six (see Figure 2).

Figure 2. Number of digital repository or digital preservation systems used at respondent institution (164 respondents)



Second, homegrown systems ranked quite high on the list: 53 of the 164 respondents to this question, or 32%, answered that they use a system that was developed in-house. This may indicate existing systems did not meet institutional requirements for a repository and/or preservation system. This interpretation is confirmed by the in-depth interviews conducted with selected survey respondents. Five of the twelve interviewees indicated that they used homegrown systems or tools in their survey responses, and in the interviews, all described various gaps the homegrown system was intended to fill. For some, a homegrown tool was developed to facilitate digital preservation activities that the current system did not do. For example, one interviewee described developing a tool to enable staff to work more easily in an existing digital asset management system. Other interviewees were motivated to develop homegrown tools because existing systems did not cover all of their needs. “Nothing else was able to deliver what we wanted,” said one interviewee, so the organization opted for a repository system that was developed in-house.

Questions for further research

Although these survey questions revealed a wealth of information, the data also brought up issues that could be the focus of further research. It’s clear that survey respondents are using many different systems and technologies, but there are many unknowns about their use. Survey respondents were not asked how long they had been using their system(s), or what version of the system(s) they used. Many repository or preservation systems have significant diversity between versions, which could influence how they are used. Similarly, the survey did not define a “homegrown system,” instead leaving it up to respondents to decide if they believed their local

tools could be categorized as such. As a result, there would certainly be a huge variety of the types of tools and technologies within this category.

Furthermore, although the survey received 170 complete responses, there were not enough responses in certain areas to analyze the data further. For example, the grant team attempted to break down the responses to Q2 ("Which of the following digital repository or digital preservation systems are used at your institution?") according to other responses, such as how many terabytes a respondent had collected or what type of institution they represented. After looking at the data through these lenses, however, the grant team decided that ultimately conclusions could not be drawn from groups as small as ten or fifteen survey respondents.

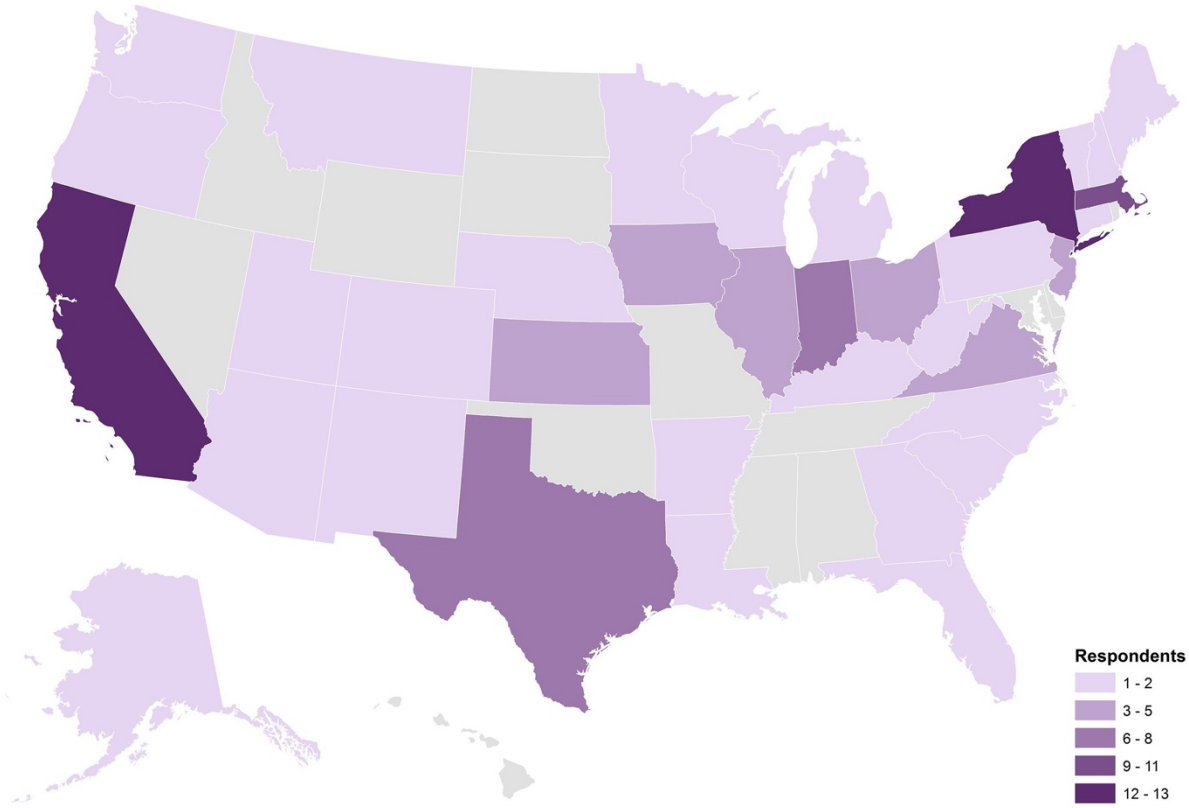
Organizational Profile

At the end of the survey, respondents were asked to provide more details about their institutions and their own roles within the organization. All survey respondents were shown these questions.

- Q24: Type of institution
- Q25: Where is your institution located?
- Q26: What best describes your role at your institution?
- Q27: Which of the following statements describes your day-to-day work? Select all that apply.

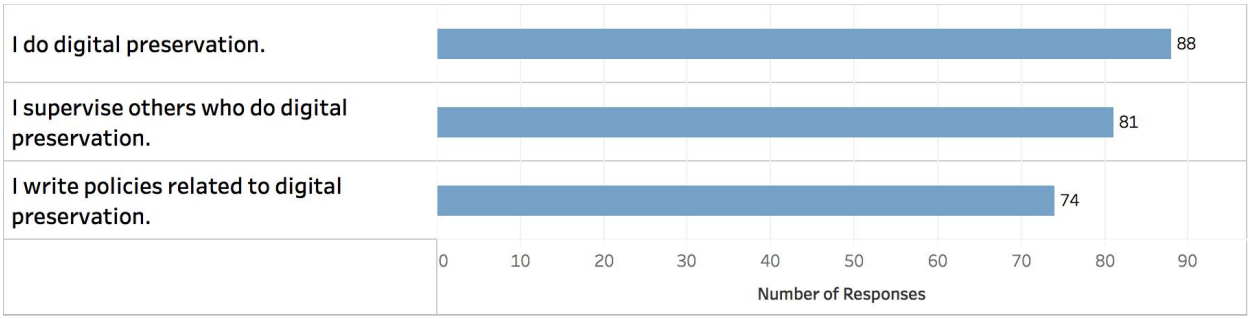
The majority of respondents (78%, or 123 of the 157 who answered Q24) self-identified as academic institutions, though representatives from archives, government organizations, museums, non-profit organizations, and public libraries also responded to the survey. Survey responses also came from a wide geographic range: of the 157 survey respondents who volunteered their location, eleven different countries are represented. The majority (82%, or 129 respondents) said they were located in the United States; these represented 35 states and the District of Columbia. Of the 28 respondents who said they were located outside of the United States, most are from Canada (11) and the United Kingdom (8). The survey also received responses from China, Denmark, Mexico, New Zealand, South Africa, Switzerland, and the Netherlands.

Figure 3. "Where is your institution located?" (157 total respondents; 28 respondents from outside the U.S. are not represented)



Survey respondents were nearly evenly split between those that identified as an administrator or department/unit head and those that identified as staff (55% versus 45%, respectively, of the 154 respondents who answered the question). Responses to question 27 reveal that many respondents are responsible for much of the digital preservation work at their organizations (see Figure 4). Of the 145 respondents who answered this question, nearly 23% (33) indicated that in their day-to-day work, they do digital preservation, write policies related to digital preservation, and supervise others who do digital preservation. About 45% (65) of the respondents to this question said their day-to-day work is characterized by more than one of those areas. Most of the respondents to this question identified themselves as digital preservation practitioners: "I do digital preservation" was selected most frequently of the three choices, at 88 times (or by nearly 61% of respondents who answered this question). Perhaps unsurprisingly, "I supervise others who do digital preservation" was chosen at about the same rate (81 times, or 56% of respondents to the question) as respondents who self-identified as administrators or department/unit heads in the previous question.

Figure 4. "Which of the following statements describes your day-to-day work? Select all that apply" (145 Respondents)



As stated above in the Methodology, the grant team selected interview subjects who represented a diverse group of institutions and digital preservation practices. The twelve interviewees represented six public university libraries, two private university libraries, two museums, one public library, and one government organization; the twelve came from nine different states in the US. The interviewees used eight different local repository or digital preservation systems between them, and were members of various DDP services including APTrust, Chronopolis, DPN, and MetaArchive. The interviewees were a diverse group in terms of their roles within their institutions and digital preservation programs as well. Many of the interviewees identified as administrators or department heads, but others identified more as staff members. In several cases, the person contacted for an interview brought additional colleagues for the conversation to present a more complete picture of the digital preservation environment at their institution. Others requested the interview questions in advance so that they could confer with colleagues prior to the interview. This demonstrates that in many cases, digital preservation is a highly collaborative endeavor.

Questions for further research

Although interviewees were asked to describe their organization's digital preservation program, including number of staff involved and their roles, it became clear that this question was more difficult to answer than the grant team had expected. Interviewees often seemed unsure whether they should mention staff whose roles contribute to digital preservation systems or workflows, such as colleagues in information technology or preservation. The grant team observed that interviewees all had different ways of talking about support from information technology departments, but generally did not have time for further questions about how peripheral staff contribute to or interact with digital preservation workflows. Future research into the collaborative and interdependent nature of digital preservation would likely be illuminating.

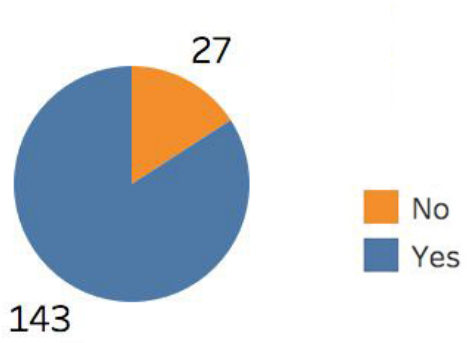
Distributed Digital Preservation

These questions were designed to gather information about distributed digital preservation practices. All survey respondents were shown Q5, but the questions that follow were dependent on the answer for Q5. Skip logic within the survey is indicated below, and a detailed visualization of the possible survey paths is found in Appendix B.

- Q5 (Required): A recommended digital preservation practice is to keep multiple copies of digital files in multiple places. Does your institution do this?
- Q8: You indicated that your institution keeps multiple copies of digital files in multiple places. How many copies (including the original file) are kept? *[Displayed if Q5 = "yes"]*
- Q9: Where are the copies stored? Select all that apply. *[Displayed if Q5 = "yes"]*
- Q10: Which distributed digital preservation service(s) does your institution use? Select all that apply. *[Displayed if respondent chose "A distributed digital preservation service (such as Chronopolis, APTrust, etc." in Q9]*

The survey data reveals that most respondents are storing copies of their unique content in multiple locations; of the 170 survey responses, 143, or 84%, indicated that they did so. Upon an affirmative answer to question 5, respondents would be prompted with questions 8 through 10. If a respondent answered negatively for question 5, he or she would instead be asked about barriers to distributed digital preservation (these questions, Q6 and Q7, will be discussed later. See Appendix B for a visualization of the question order). The number of copies stored varied among respondents, however.

Figure 5. "A recommended digital preservation practice is to keep multiple copies of digital files in multiple places. Does your institution do this?" (170 respondents; all survey respondents required to answer)

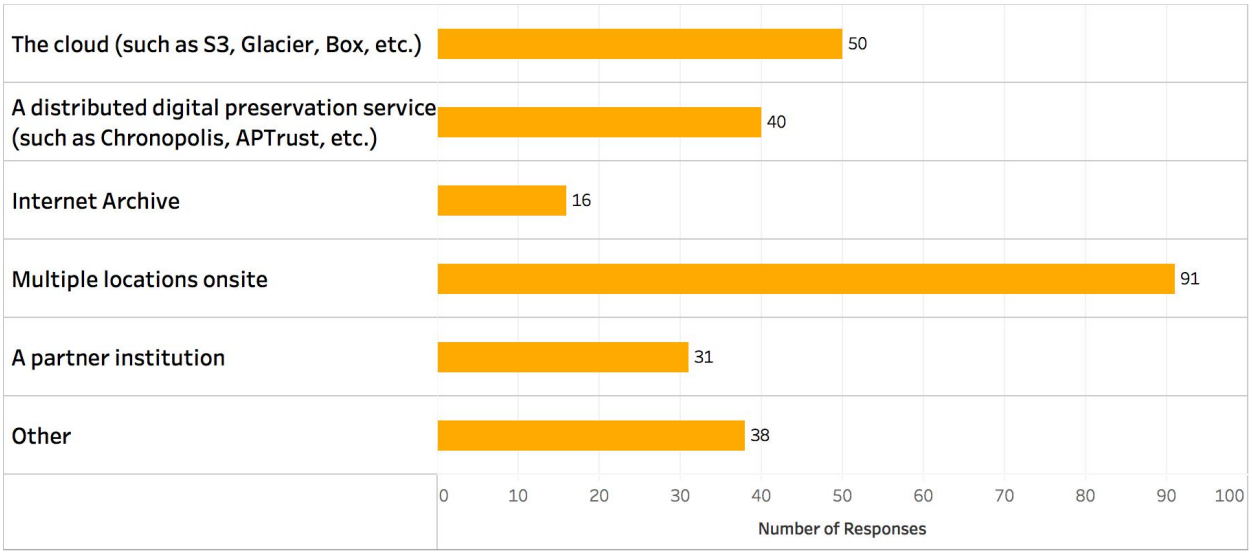


Of the 130 respondents who answered the question on how many copies they have, the majority - 74, or 57% - said three. Storing two copies was reported by nearly a quarter (30) of those who answered this question. Very few respondents said that they store five or six copies, though seven and above is more common: ten respondents reported keeping seven or more copies. All ten who indicated that they store seven or more copies also reported being members of a private LOCKSS network (PLN). However, the opposite is not true: not all survey respondents who reported being a member of a PLN also indicated that they keep seven or more copies.

The survey posed two further questions to respondents who reported that they store copies of their data in multiple locations: where the copies are stored (Q9), and if the respondent selected a distributed digital preservation system, which one (Q10). As with the question on systems

(Q?), respondents were able to choose multiple answers to indicate where they store copies. Q9 received 266 responses overall (for 137 who answered the question), indicating that many respondents are pursuing multiple storage strategies. Answers to this question demonstrate that the most common storage strategy is overwhelmingly to keep copies onsite (see Figure 6).

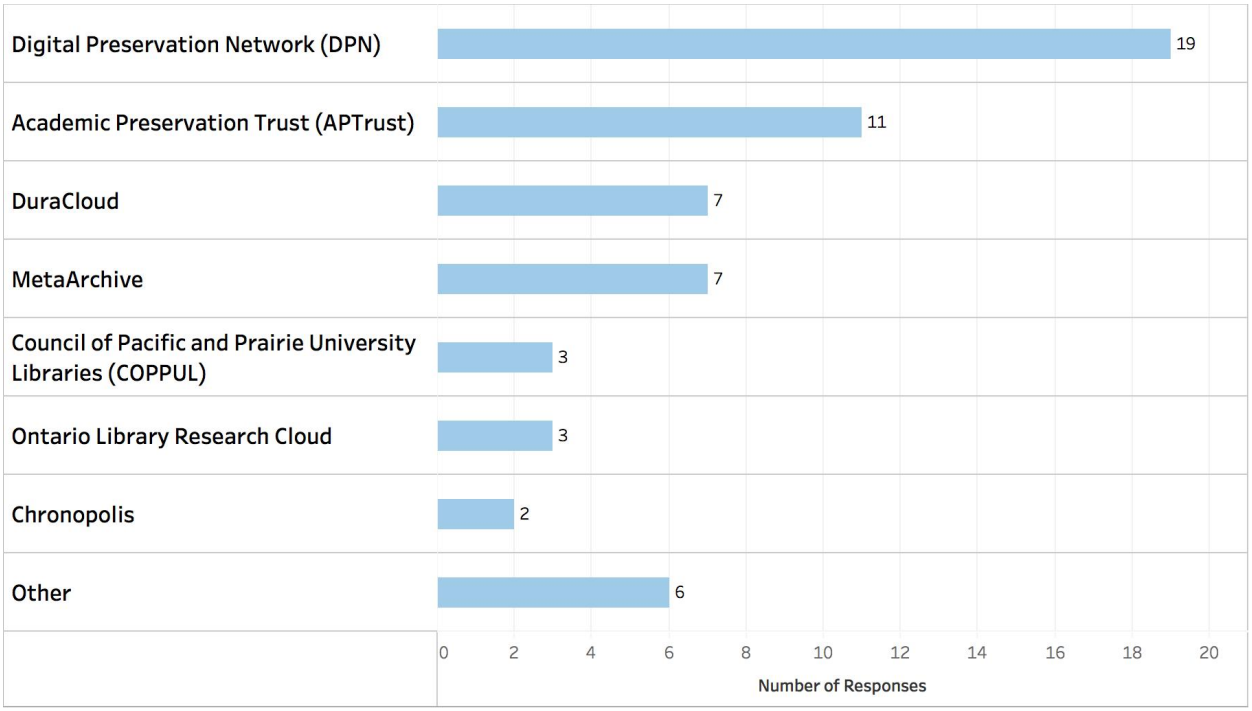
Figure 6. "Where are the copies stored? Select all that apply" (137 respondents)



Ninety-one of the 137 respondents indicated that they keep copies in multiple locations onsite. Eight respondents chose the "other" option to express that they have only one copy onsite (in addition to at least one other storage strategy, like the cloud or a DDP service). When these are combined with those who keep multiple copies onsite, the number of respondents storing either one copy onsite or multiple copies onsite rises to 72% of those who answered Q9. Cloud storage and distributed digital preservation (DDP) services are also popular storage locations, selected by 50 and 40, or 36% and 29% respectively, of respondents to this question.

All of the forty respondents who indicated that they use a DDP system answered Q10. This question also allowed respondents to select more than one option, and resulted in 58 responses for the forty who answered (see Figure 7).

Figure 7. "Which distributed digital preservation service(s) does your institution use? Select all that apply" (40 respondents)



The data reveal several points of overlap in DDP system use. First, although the Digital Preservation Network (DPN) was selected most frequently at nineteen times (47% of respondents to the question), five respondents indicated that they use DPN in conjunction with Academic Preservation Trust (APTrust) or Chronopolis. Respondents may have selected both because APTrust and Chronopolis are nodes of DPN, not because they are members of both DDP systems separately. Second, all seven respondents who indicated that they use DuraCloud also selected another DDP system. Discussions with interviewees on this topic indicate that DuraCloud is frequently used as an interface to other systems. As one interviewee described their workflow: "With DPN we use the DuraCloud interface, and then [the data] gets transported to Chronopolis at the point it's finished."

Questions for further research

The grant team was surprised by some of the data on storage; specifically, the grant team expected to see a higher rate of DDP use. Future research in the area of DDP use may want to investigate why more institutions are not utilizing DDP services. Respondents to this survey were not asked this directly,[4] and although interviewees were asked why their institutions had chosen specific storage strategies, only twelve survey respondents were interviewed so conclusions cannot be drawn about general reasons for selecting certain storage strategies.

[4] Survey respondents who indicated that they do not keep multiple copies of digital files in multiple locations were asked why later in the survey (discussed below), but this question was asked of *all* respondents who said they do not keep multiple copies, not only those who said they do not use a DDP service.

Tracking

If they had indicated that they keep multiple copies of digital files in multiple locations, survey respondents were also asked how they keep track of these copies.

- Q11: How do you or your colleagues keep track of the copies that are stored in different places? *[Displayed if Q5 = "yes"]*

Answers to this question were collected via a free text field; 102 of the 143 survey respondents who answered "yes" to Q5 answered Q11. These responses indicated a variety of methods for tracking copies which were then categorized by the grant team as:

- Manual tools (31 respondents): spreadsheets, databases, catalog records
- Automated tools (30 respondents): logs, backup reports, third-party services
- Tools provided by DDP systems: (11 respondents): MetaArchive's Conspectus tool, DuraCloud content manifests, APTrust dashboard
- Don't keep track (11 respondents)
- IT support (8 respondents): monitoring, backups, reports
- Homegrown tools (5 respondents): inventory software

The issue of tracking copies of content between systems was an obvious pain point for both survey respondents and interviewees. It was also mentioned by eleven percent (14 out of 127) of respondents to Q23 ("What is lacking/missing in the technology being used at your institution?" Q23 is discussed in greater detail below). Several interviewees described their manual method of tracking their copies:

> *"We've got multiple people who are pushing this data so we're just essentially making a spreadsheet of what has been pushed where so that we know and so that we don't duplicate the labor. Sounds a little crazy but I don't know any other way to do it because we don't have any sort of preservation management system. Like we have no tracking tool."*

Others had access to automated tools or tools provided by DDP systems but also indicated that they did not rely solely on those tools.

> *"And then how I track what I send to [DDP service], that is all on a spreadsheet. [DDP service] does maintain this separate web application...to view what they have in the network, how many copies there are, you can run some fixity verification reports, and things like that. But I kind of rely more so on that spreadsheet."*

Curating Distributed Materials

Factors affecting curation decisions

Storing additional copies of materials off-site usually incurs costs, which generally increase as the amount of data increases. In order to uncover curation practices surrounding sending subsets of materials off-site, the survey posed the following questions:

- Q17: Earlier you indicated that your institution keeps multiple copies of digital content. Does your institution select a subset of your data to go to a distributed repository (or offsite, or the cloud)? *[Displayed if Q5 = "yes"]*
- Q18: What proportion of material is distributed to the following systems? *[Displayed if Q5 = "yes." Selections from Q9: "Where are the copies stored?" carried forward and displayed as the systems]*
- Q19: Please drag and drop to rank the criteria used to determine which material go to distributed locations. *[Displayed if Q5 = "yes" and Q17 = "yes"]*

Almost half of the survey respondents who responded "yes" to keeping multiple copies in multiple places in Q5 (49%, or 64 of 130 responses) indicated that they sent a subset of their data to a distributed repository (or offsite, or to the cloud) in Q17. The respondents that indicated a subset of data was sent to a distributed repository were then asked to quantify the amount of data sent in Q18 to the various options they had selected from Q9 ("Where are the copies stored?"). Those storage options included:

- The cloud (such as S3, Glacier, Box, etc.)
- A distributed digital preservation service (such as Chronopolis, APTrust, or similar service)
- Internet Archive
- A partner institution
- Multiple locations onsite
- Other (list all):

Percentages totaling above 100% were allowed to account for content placed in multiple storage configurations. Sixty-two respondents reported that they stored 95% or more of their content in multiple locations onsite, and twenty-four stated they stored the same percentage in the cloud. For those respondents who reported storing data in distributed digital preservation systems, 17 kept 1-25% of their data in those systems, 9 kept 30-90%, and 14 kept 95% or more. The majority of the group keeping 95% or more in distributed digital preservation systems (12 of 14) reported that they had less than 50 terabytes of preservation materials in Q3.

Figure 8. "What proportion of material is distributed to the following systems?" (120 respondents)

	1 - 25%	30 - 90%	95% or more
The cloud (such as S3, Glacier, Box, etc.)	5	8	24
A distributed digital preservation service (such as Chronopolis, APTrust, or similar service)	17	9	14
Internet Archive	11	3	1
A partner institution	3	9	12
Multiple locations onsite	1	13	62
Other	4	3	16

When these respondents were asked to rank the importance of criteria used to select the subset of materials sent off-site, the majority chose *Mandate (legal, grant, or other)* (17), followed closely by *Intrinsic value* (15) and *Content type* (10) as the most important considerations. *Intrinsic value* (16) was a popular second pick, as was the *Preservation state of the original* (12). *Access restrictions*, *Rights restrictions*, the *Cost to acquire the materials*, and the *Risk to reputation* were all rated as low in importance by survey respondents.

Figure 9. "Please drag and drop to rank the criteria used to determine which material go to distributed locations" (58 respondents)

Most important ↓ Least important

Ranking	Access restrictions	Content type (such as audiovisual, text, etc.)	Cost to acquire	Cost to digitize	Intrinsic value	Mandate (legal, grant, or other)	Preservation state of original (physical) item	Rights restrictions	Risk to reputation
1	2	10	3	3	15	17	4	2	2
2	1	6	4	4	16	9	12	4	2
3	8	6	3	8	7	7	10	5	4
4	2	11	5	9	8	5	6	7	5
5	9	8	3	7	3	2	8	6	12
6	8	6	9	6	3	8	8	4	6
7	12	6	9	9	4		1	12	5
8	7	2	10	8	1	3	5	12	10
9	9	3	12	4	1	7	4	6	12

Use and influence of policies on curation decisions

- Q20: Does your institution have preservation policies that help guide your selection decisions for materials held locally?
- Q21: Does your institution have preservation policies that help guide your selection decisions for materials going to distributed systems? *[Displayed if Q5 = "yes"]*

In addition to identifying which criteria was used in determining the subset of materials to be sent to a distributed repository (or otherwise offsite), the survey asked if there were institutional policies in place to help guide selection decisions for materials held locally. Of 156 respondents, the majority (93, or 60%) responded in the affirmative. The same question was asked regarding selection decisions guiding materials sent to distributed systems (Q21), and of the 129 respondents, 61 (47%) replied yes. Q21 was only shown to respondents that had replied that they did keep multiple copies of digital files in multiple places in Q5.

The in-person interviews provided additional insight into the selection processes institutions use to send subsets of their materials off-site. Four interviewees mentioned grouping their preservation materials into tiers based on various qualities of the materials: their uniqueness, institutional value, born digital materials vs. digitized, and the format of the original for digitized materials (ie. photographs versus audio-visual materials).

> *"… it boils down to: is it born digital? Is it something that has preservation needs? So if we scan something that is super fragile, or that is in an unstable format, or if it's something that we pay for a vendor to reformat, like a reel to reel tape or something, those are all going to be things we do full-fledged, full-on digital preservation for. But if it's something that we're just scanning, like if it's a collection of letters than we're just scanning to make available online for researchers, we're not going to do full on digital preservation because it's too expensive."*

Of these interviews, one only institution had codified these curation decision factors into policy and two interviewees indicated that they were working with draft versions of a policy. All four interviewees linked these preservation tiers to corresponding levels of preservation care for the materials. The higher the risk of data loss, the higher the level of preservation the materials would receive. Distributed digital preservation systems were considered the highest level of preservation. Cloud storage was considered a lower-level of preservation, to be used for materials appraised at lower tiers.

> *"If there is more of a likelihood that we could lose that content ... well if the risk involved in losing that content would mean losing that information entirely, then I want that content to be preserved in MetaArchive."*

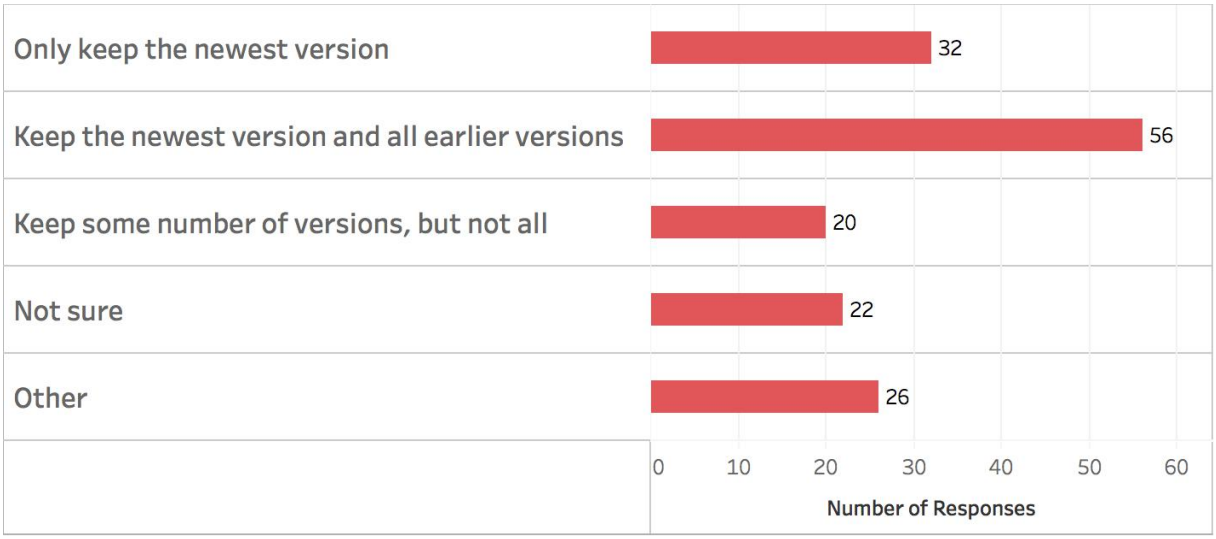
Versioning

The survey asked a series of questions to uncover information about how institutions are handling the versioning of both content files and their associated metadata, in both their local repositories and distributed copies.

- Q12: When updates or changes occur to the preserved content, how do you version the data stored in your *local* repository or preservation system?
- Q13: When updates or changes occur to the preserved content, how do you version the copies of your data that are stored in *distributed* systems? *[Displayed if Q5 = "yes"]*
- Q14: Does your institution version metadata separately from content? *[Displayed if Q5 = "yes"]*
- Q15: When updates or changes occur to the metadata for preserved content, how do you version the metadata? *[Displayed if Q5 = "yes" and Q14 = "yes"]*

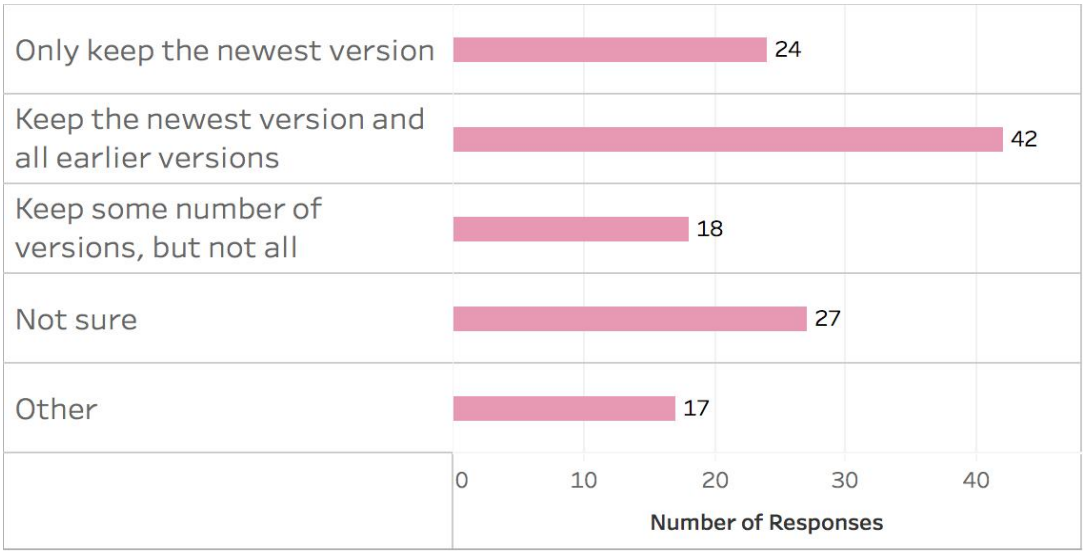
Responses for Q12 and Q13 had similar distribution patterns, with the majority reporting that they keep all versions of the data (36% and 33%, respectively). Respondents that selected "Other" were asked to describe their versioning practices. For Q12, the majority of these "Other" descriptions (13 of 26) were a variation of "it depends," with the dependencies often being related to the type of content or the system it is in.

Figure 10. When updates or changes occur to the preserved content, how do you version the data stored in your local *repository or preservation system? (156 responses)*



Q13 was only displayed to respondents who answered affirmatively to Q5. The percentage of respondents reporting that they were "Not sure" of their versioning practices increased from 14% (22 responses) in Q12 to 21% (27 responses) in Q13. Of the "Other" descriptions provided for Q13, the majority (6 of 17) stated that they have not versioned the copies stored in distributed systems.

Figure 11. When updates or changes occur to the preserved content, how do you version the copies of your data that are stored in distributed *systems? (128 responses)*



When asked if institutions versioned metadata separately from content (Q14), respondents split almost in half, with 41% (53 of 130) responding “yes” and 43% (56) responding “no.” Sixteen percent (21) responded that they were unsure. When asked which versions are kept for metadata in Q15, 37% (19 of 52) stated that they kept all versions of the metadata.

When asked about versioning practices, interviewees noted the distinction between receiving versioned materials (notably special collections) and versions of digitized materials:

> *“...We actually assume like everybody else we have to support versioning... Then we realized, wait a second, what are the real use cases? For special collection stuff, all versions are first class objects. That's almost the whole point of some of these things. If somebody has five manuscripts of somebody's book, well they're all equally important. There's no question of versioning. On the digital size, the respective digitization side, it's way too expensive to go back and produce another TIFF, if that ever happens, because the first group of TIFFs were a mistake, you don't want to keep them. For digital archives, versioning is moot. For digital collections, it's too expensive.”*

> *... By the time it gets to the repository and they say, "Okay, we're ready to ingest this collection." The images are assumed to be in their final states. If there's ever a need to replace an image ... Sometimes things need to be flipped or read more or whatever…. In those cases, we simply just replace the previous image.*

Not every organization reported replacing the old master digitized versions with new or corrected digitized versions; more than one organization kept all versions of the data. Many weren’t sure how many versions were kept because versioning decisions had been made in the past and were not communicated to the current system administrators, or the determination was based on the systems used:

> *"I think we have some old formula that keeps, this is again in the homegrown system, that keeps like the first version and then the most recent percentage three or four if we update something, but that's it… We're doing something, and we're doing something based on a historical decision years ago."*

> *"Our versioning varies by repository application essentially. DSpace for example, we don't have any versioning. For our Fedora and our Fedora Hydra-based repository applications, it varies. In some, we do have versioning of some parts of the objects. It's essentially whether versioning is turned on in Fedora, so we are doing versioning in some cases and not in others. I think our versioning has been somewhat haphazard than deliberate... They were decisions that were made for reasons at the time, but those reasons aren't always clear."*

Metadata versioning practices also ran the gamut from no versioning to full versioning:

> *"Basically right now, we don't version anything because we don't have good use cases, and that includes metadata. If somebody gives us a new file, it's because the first one was wrong."*

> *"We felt like we wanted to be able to maintain versions of a metadata just to track any of those changes. Also because turning that on is really simple to do in Fedora. With metadata, you're not taking up a lot more space by doing that, so it just seem like ... We're basically getting this for free, so why not?"*

Some interviewees did not manage metadata separately from content data:

> *"Our EAD finding aids and our digital objects are stored together in the AIP in our current preservation system. Let's say that we need to replace the finding aid because we made an edit or something. Our programmer will manually rename that file and plop the new one in and re-send that, and accept. It's this real manual versioning going on, but it's not really even true versioning. It's not recording exactly what was changed…. "*

Interviewees did envision future systems that would provide the versioning capacities they desired:

> *"And we have all sorts of plans to build a versioning system that will allow us to version both metadata and digital objects themselves, or like the package of digital objects. But we just haven't built it yet."*

> *"I think our versioning is not what it could be…. we're looking at things like object store systems so that versions are kind of automatically kept track of since we can automatically determine, not reproduce duplicates.... I don't know if you look at the way that Center for Open Science does forking of data or they do forking of entire projects that are sets of files, and I think it's as model that's sort of based on GitHub probably where you're keeping track of changes in a pretty efficient way. We look at things like distributed file services like IPFS, like we're tracking that pretty carefully…. Not sure when we want to*

> *jump into that pool, but those kinds of things seem like they're sensible approaches to thinking about versioning."*

Interoperability

- Q23: What is lacking/missing in the technology being used at your institution?

Twenty-six respondents to Q23 identified the overspecialization of their systems as a contributor to interoperability issues. One respondent wrote that what is missing from their systems is "the glue to hold these silos together." Twenty-one responses to the question were characterized by the grant team as a lack of integration between systems.

When asked about interoperability and communication between preservation systems, most of the interviewees expressed a lot of frustration that their systems and tools did not work together more effectively: "Right now, nothing is actually interacting together."

> *"I think interoperability itself is the main challenge that we're facing, to be able to get these different systems to work together, whether it's our descriptive systems or preservation."*

> *"Wouldn't it be wonderful if we had some sort of management tool that could talk to all of these places and if you just put content in this one little spot, all of these places could just pull from that, and it could be fully automated."*

Packaging

The Open Archival Information System (OAIS) Reference Model defines three types of information packages for logical containers of content and optional descriptive information: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP). The general workflow defined by the model involves SIPs generated by information producers and sent to an OAIS that generates AIPs from one or more SIPs. The OAIS is then responsible for handling the transformation of one or more AIPs into DIPs for consumer access (CCSDS Secretariat, 2012, p. 2-8). DDP systems typically support packages generated from using the BagIt specification which is used to generate "bags" containing directories of files along with relevant metadata. Off-the-shelf tools such as Bagger, created by the Library of Congress, are often used to generate SIPs for ingest in DDP systems. However, survey respondents and interviewees indicated that homegrown tools and systems are being used to generate SIPs as well. The OAIS model came up frequently during the interviews, with eight out of the twelve interviewees mentioning some of the OAIS model typically related to packaging.

Several interviewees described a manual process for generating bags outside the context of their local repository systems. These processes typically involve someone organizing files into a directory structure according to a predefined naming scheme, and then packaging those

directories of files into bags using a tool such as Bagger. Interviewees were asked whether they had any notable workarounds in their workflows, and two referenced “off-the-shelf” tools like Bagger as an example to why workarounds were less necessary now. One interviewee mentioned extending the functionality of Bagger as a form of workaround, because they wanted more flexibility in terms of compression and output file types from the tool.

One interviewee described a desire for bags that are packaged in a way that makes them useful outside of repository in which they were generated: “I don't want something going in [to the DDP] that requires Fedora to interpret it, as I want to have packages that stand on their own and can understand outside the context of Fedora.”

Data restoration

The survey did not address the topic of data restoration or recovery, but the question was asked in the follow-up interviews. Three interviewees noted that in the event of data loss in their local repository, recovery from the bags stored in their DDP would require some degree of processing. A few examples of the type of processing described are parsing metadata for many files that had been collocated into one file stored in a bag, or matching files and metadata that were stored in different directories within bags. When discussing data recovery from their DDP, one interviewee noted, “I think on the restore side, if you have everything in bags, it shouldn't be too much of a surprise what you get.”

Organizational Challenges

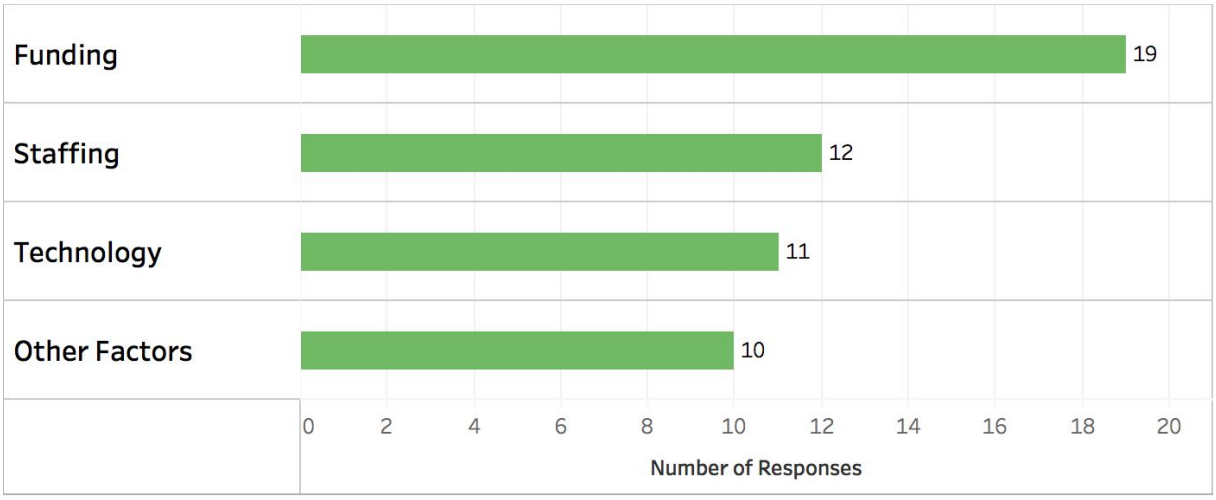
Barriers to Distributed Digital Preservation

There were several other notable trends in survey responses and the interviews. In addition to discovering common digital preservation practices, the survey was also designed to reveal barriers. If survey respondents answered “no” to question five, indicating that they did not keep multiple copies of their content in multiple locations, they were prompted with a follow-up question asking them to describe reasons why.

- Q6: Why doesn’t your institution keep copies in multiple locations? What are the barriers to doing so? Select all that apply.
- Q7: For your choices above, please explain the barriers in more detail.

Of the 27 survey respondents who said that they did not keep multiple copies of data in multiple locations, 26 chose to describe why. Respondents were prompted to select all that applied from “funding,” “staffing,” “technology,” or “other factors,” and were given the option to further explain if they chose the latter. Their responses indicate that there are often multiple obstacles to storing more than one copy: sixteen of the 26 respondents to this question chose more than one reason. Funding (19) was the most common barrier cited, but respondents chose the remaining options in virtually equal numbers (see Figure 12).

Figure 12. "Why doesn't your institution keep copies in multiple locations? What are the barriers to doing so? Select all that apply" (26 respondents)



Most of the ten responses of "other factors" as barriers to storing multiple copies in multiple locations can be characterized as either institutional barriers, or respondents stating that they are still in the planning stages of digital preservation. For example, one respondent noted that, "Despite advocacy, we do not completely have a collective vision yet in terms of preservation." Five others also described their barriers as institutional roadblocks. The seven respondents who described their digital preservation program as being in early stages generally outlined aspirational plans towards keeping multiple copies of their content in multiple locations. "We are currently investigating preservation systems and workflows in hope to soon be compliant with this recommendation," said one respondent.

Barriers to Adopting Digital Preservation Policies

Similarly, if survey respondents answered "no" to either question 20 or 21, indicating that they did not have preservation policies for either locally stored materials or those stored in DDP systems, they were prompted with a follow-up question asking them to describe the barriers to adopting preservation policies.

- Q22: What are the barriers to adopting preservation policies at your institution?

Like the responses regarding obstacles to storing multiple copies in multiple locations, survey responses reveal that there are often many barriers to adopting preservation policies. Thirty-four respondents (of the 78 who chose to answer the question) described the barriers they face with multiple reasons, such as "staff, time, money, expertise." The grant team coded the responses to this question by grouping responses into common themes, and the most frequent reasons given were coded as staffing problems (22) and lack of organizational coordination (20). As described by respondents, the latter usually manifests as lack of consensus across units, or a decentralized structure that makes it difficult to develop or implement organization-wide policies.

The barriers cited by survey respondents were also echoed by interviewees whose organizations did not have preservation policies for digital content. One interviewee described being part of a digital preservation policy committee several years ago, but the committee "...basically lost its momentum" before developing any policies. Another interviewee, noting that they have struggled with lack of administrative support for digital preservation, stated, "What's the point of putting a policy together if there's absolutely no capacity to actually do anything about it?"

Funding & Staffing

Lack of staff or funding emerged as a frequent theme throughout the survey and interviews. Not only did survey respondents cite both as barriers to distributed digital preservation and adopting digital preservation policies, but they also frequently mentioned staff or funding when asked what is lacking or missing from the technology used at their organization (Q23). One survey respondent described organizational barriers in this way:

> *"Convincing library management to allocate sufficient resources for digital preservation has been extremely difficult. There have been multiple efforts and strategies over the years, but we do not have sufficient in-house technical staff or available operating funds."*

Insufficient staffing or funds were also discussed by interviewees in many different contexts. When asked why their organization had chosen a particular digital preservation system or set of systems, interviewees frequently mentioned weighing the resources -- in terms of both cost and staff time -- that would be required for various systems. The lack of staff, or staff turnover in key roles, was the organizational challenge most frequently experienced by interviewees. Nearly all of the interviewees mentioned struggling with shortcomings in staffing in some way. Three interviewees described facing hiring freezes, or long periods of time in which key staff members who had left were not replaced. For example, one interviewee described her struggle to replace a key role in her organization's digital preservation program: "if somebody leaves, they leave the position open for a year to see whether you really need it...before they'll approve a hire or replacement."

Interviewees also frequently cited lack of organizational change in terms of staffing as a significant obstacle. After describing how various digital preservation tasks are distributed among several staff members, one interviewee noted, "For none of us is this a primary duty. This is all an additional add-on." Another said that she struggled to keep up with the challenges of building a robust digital preservation program as her role within the organization kept changing and she was given more and more responsibilities.

Staff turnover was another obstacle interviewees frequently mentioned as a significant impediment to digital preservation in their organizations. Several interviewees struggled to describe some aspects of their digital preservation programs, explaining that recently departed colleagues were the primary staff members with these responsibilities. Many of the interviewees described struggling with staff retention, especially of developers and other technical staff. As one noted, "It's a pretty high-demand series of talents that we're looking for, and we can't retain

these people." Others expressed frustration that their organizations could not compete with salaries offered for technical roles in the private sector.

The survey responses and interviews clearly indicate that lack of funding and staffing are significant impediments to building a robust digital preservation program. Although the research for this grant focused on challenges in integrating digital preservation systems and technical obstacles, survey respondents and interviewees made it clear that problems acquiring and retaining adequate funds and skilled staff are considerable barriers to successful digital preservation.

Questions for further research
Further research into these issues could investigate the budgetary trends behind digital preservation programs. For example, is the funding for digital preservation activities proportional to the overall budget of the organization? What effects do overall budget cuts have on digital preservation policies and workflows? In addition, it may be interesting to investigate whether respondents feel that digital preservation is part of the mission of their institution, and whether this affects the success of digital preservation efforts.

Recommendations for Technical Solutions

The grant team and the advisory board have coalesced around three recommendations after reflecting on the survey and interview results. It is our hope that these recommendations are considered for any follow-on work from this project with the aim of improving DDP workflows and interoperability.

Recommendation 1: Develop a decision-making toolkit for choosing materials to send to DDP systems. A toolkit can help walk end-users through a series of predetermined steps in order to streamline their workflows. Survey results showed a number of reasons that people choose to send material to DDP systems, e.g. mandate, intrinsic value, etc. This work would be comparable to the MediaSCORE & MediaRIVERS projects[5] developed by AVPreserve and Indiana University for making prioritization decisions regarding digitization efforts. An analogous "Preservation Triage Tool" could go a long way towards helping those tasked with gathering resources for ingest into DDP systems to make decisions about which materials to select. Perhaps a follow-on group could decide upon the criteria in narrative form, which could be delivered as a specification that various organizations could then decide to develop tools or products for internal or shared use.

Recommendation 2: Determine a shared BagIt profile for DDP systems. This work would entail choosing a small set of fields to include in bag-info.txt that would be supported by all DDP providers. These fields could include data that would be useful across any DDP system such as internal identifier, external identifier (DOI/ARK, etc.), date created, date modified, etc. The focus of this recommendation is increased interoperability between DDP systems.

[5] https://www.avpreserve.com/mediascore-mediarivers/

Recommendation 3: Interviewees mentioned the desire for a dashboard provided by DDP systems, that would enable them to audit which resources have been distributed and possibly correlate this with additional information such as fixity checking. DDP providers could populate a dashboard with data extracted from these elements in the DDP BagIt profile, and could allow shared development among DDP systems and interested parties that could provide development support.

User Stories

The following personas are genericized examples derived from the institutions and roles of survey respondents and interviewees. Each user story contains a description of the role and responsibilities relevant to digital preservation. The effects of our technical recommendations are also described to give a sense of their potential benefits for each role.

Director of Digital Services, Private Research University Library

Role and responsibilities:

- Strategic planning for digital preservation initiatives, setting goals for local and distributed digital preservation systems
- Supervises team comprised of a digital archivist, two programmers, and four support staff
- Coordinates with Information Technologies for infrastructure and storage services
- Partners with Scholarly Communication who manage Institutional Repository service
- Institution manages 500TB of content

Effect of Technical Recommendations:

- Recommendation 1: The toolkit would provide a framework for discussing strategies with staff. It could potentially raise awareness within the organization about the variety and value of collections materials.

Digital Preservation Librarian, Large Public University Library

Role and responsibilities:

- Develops workflows for managing digitized and born-digital collections materials.
- Liaises with Special Collections curators
- Identifies material stored in local repository to be distributed to DDP services
- Works with developers to create tools for bagging and depositing content to DDP services
- Institution manages 2PB of content, mostly digitized video collections

Effect of Technical Recommendations:

- Recommendation 1: Uses the toolkit for guided selection of materials to be sent to DDP systems.
- Recommendation 3: Tracking content distributed in DDP system is made easier through a dashboard built using the shared profile, which makes it easier than manually updating spreadsheets to track which local content has been sent.

Digital Repository Specialist, Art Museum

Role and responsibilities:

- Liaises with curators and registrars
- Prepares material for ingest
- Quality control from submission to storage
- Issues support tickets with vendors
- Institution manages 10TB of content

Effect of Technical Recommendations:

- Recommendation 1: Populates data in the toolkit for those charged with making DDP material selection.
- Recommendation 2: Uses shared BagIt profile to prepare and store material across systems

Conclusion

To return back to the research questions which served as the impetus for the survey and interviews, the grant team found that although some questions were clarified, others were unveiled to be more complex and multi-layered than expected. When curating objects to ingest into a long-term dark preservation system, mandate and the intrinsic value of the materials are regarded as important, but the interviews also unsurfaced a more intricate valuation system. This system of curation criteria is one that is often not codified into official policy or procedure, but that ranks digital materials into different tiers worthy of varying levels of preservation, the highest tier being a distributed digital preservation system.

Versioning of object and metadata was a particularly murky area that was not fully clarified in this research. Survey respondents reported keeping all versions of materials, but the interviewees often relayed a different story. Interviewees often reported that they were either unsure of their versioning practices, that those practices varied depending on the materials, and/or that those practices were system-dependent. Questions about why versioning was important also arose - interviewees recognized the importance of tracking changes to the materials and corresponding metadata over time, but also the futility of keeping erroneous scans or tracking incorrect metadata, and also that special collections materials generally are treated as objects in and of themselves.

The variety of systems used and workarounds needed to use those systems effectively made it difficult to discern what technological solutions could make systems more interoperable; however, the prevalent use of bags for data packaging became evident in the interviews. Bags have become a de facto standard for packaging preservation data and are also commonly used in distributed digital preservation systems. Increasing the interoperability of bags between systems could increase overall system interoperability.

Overall, the “Beyond the Repository” planning grant project uncovered some significant and valuable findings on the integration of local digital preservation practices with distributed digital preservation. The grant team was able to recommend three solutions that will promote interoperability between local and distributed systems: a toolkit for making curation decisions, a shared BagIt profile for DDP systems, and a dashboard that would allow users to track or manage content they have sent to DDP systems. However, it is clear that more research needs to be done in this area, not only to dig deeper into research questions addressed in this grant, but to keep pace with the developing field of distributed digital preservation as it inevitably changes and evolves.

Appendix A: Survey Questions

Q1: The aim of this survey is to better understand the challenges surrounding the integration of local digital preservation services with distributed preservation networks. Data collected from this survey will help identify and inform broadly- applicable solutions to these challenges.

This survey is intended for organizations who have committed to long- term preservation of their digital assets, whether that is fulfilled in-house or outsourced to a commercial, nonprofit, or consortial provider. We encourage participation from all types of organizations.

We will make our best effort to protect your individual survey responses so that no one will be able to connect your responses with you or your organization. Any personal information that could identify you or your organization will be removed or changed before results are made public. We will combine your responses with the responses of others and make the aggregated results public, and preserve the anonymous data long- term for research purposes. At the end of the survey you will be asked to provide your contact information for the project team to follow up with you on your institution's digital preservation practices. Any identifying information gathered for these voluntary follow--up interviews will also be anonymized before results are made public.

This project was made possible in part by the Institute of Museum and Library Services, Grant LG- 72--16--0135--16.

Q2: Which of the following digital repository or digital preservation systems are used at your institution? Select all that apply. [Options included: Archivematica, BePress, ContentDM, DSpace, DuraCloud, EPrints, Fedora, Hydra, Homegrown system, Islandora, Preservica, Rosetta, Other (list all)]

Q3: Approximately, how many terabytes of unique content has your institution collected? [Options included: Less than 1 terabyte, 1-10 terabytes, 11-50 terabytes, 51-100 terabytes, More than 100 terabytes]

Q4: In the next year, how much *additional* data do you expect your institution to collect? [Options included: Less than 1 terabyte, 1-10 terabytes, 11-50 terabytes, 51-100 terabytes, More than 100 terabytes]

Q5 (Required): A recommended digital preservation practice is to keep multiple copies of digital files in multiple places. Does your institution do this? [Options included: Yes, No]

Q6 [provided as follow up if “No” was selected in Q5]: Why doesn’t your institution keep copies in multiple locations? What are the barriers to doing so? Select all that apply. [Options included: Funding, Staffing, Technology, Other factors]

Q7: For your choices above, please describe the barriers in more detail.

Q8 [displayed if “Yes” was selected in Q5]: You indicated that your institution keeps multiple copies of digital files in multiple places. How many copies (including the original file) are kept? [Options included: 2, 3, 4, 5, 6, 7+]

Q9: Where are the copies stored? Select all that apply. [Options included: The cloud (such as S3, Glacier, Box, etc.), A distributed digital preservation service (such as Chronopolis, APTrust, or similar service), Internet Archive, A partner institution, Multiple locations onsite, Other (list all)]

Q10: Which distributed digital preservation service(s) does your institution use? Select all that apply. [Options included: Academic Preservation Trust (APTrust), Alabama Digital Preservation Network (ADPNet), Chronopolis, Council of Pacific and Prairie University Libraries (COPPUL), Digital Preservation Network (DPN), DuraCloud, MetaArchive, Ontario Library Research Cloud, Persistent Digital Archives and Library System (PeDALS), Other (list all)

Q11: How do you or your colleagues keep track of the copies that are stored in different places?

Q12: When updates or changes occur to the preserved content, how do you version the data stored in your **local** repository or preservation system? [Options included: Only keep the newest version, Keep the newest version and all earlier versions, Keep some number of versions, but not all, I’m not sure, Other (please describe)]

Q13: When updates or changes occur to the preserved content, how do you version the copies of your data that are stored in **distributed systems**? [Options included: Only keep the newest version, Keep the newest version and all earlier versions, Keep some number of versions, but not all, I’m not sure, Other (please describe)]

Q14: Does your institution version metadata separately from content? [Options included: Yes, No, I’m not sure]

Q15: When updates or changes occur to the metadata for preserved content, how do you version the metadata? [Options included: Only keep the newest version, Keep the newest version and all earlier versions, Keep some number of versions, but not all, I’m not sure, Other (please describe)]

Q16: A preservation event is an action that involves or impacts at least one object (or agent) associated with or known by the preservation repository. Examples include capturing information about when an object is ingested into a preservation repository, when it is versioned, etc. Does your institution log preservation events? [Options included: No, we don’t log preservation events; Yes, my system does it for me; Yes, but it’s not automated (please describe)]

Q17: Earlier you indicated that your institution keeps multiple copies of digital content. Does your institution select a subset of your data to go to a distributed repository (or offsite, or the cloud)? [Options included: Yes, No]

Q18 [provided as follow up if “Yes” was selected in Q17]: What proportion of material is distributed to the following systems? Note: the total can add up to more than 100. [Options included: The cloud (such as S3, Glacier, Box, etc.), A distributed digital preservation service (such as Chronopolis, APTrust, or similar service), Internet Archive, A partner institution, Multiple locations onsite, Other (list all)]

Q19: Please drag and drop to rank the criteria used to determine which material go to distributed locations. [Options included: Risk to reputation, Cost to acquire, Access restrictions, Preservation state of original (physical) item, Rights restrictions, Mandate (legal, grant, or other), Cost to digitize, Content type (such as audiovisual, text, etc.), Intrinsic value]

Q20: Does your institution have preservation policies that help guide your selection decisions for materials held **locally**? [Options included: Yes, No]

Q21: Does your institution have preservation policies that help guide your selection decisions for materials going to **distributed systems**? [Options included: Yes, No]

Q22 [provided as follow up if “No” was selected for either Q20 or Q21]: What are the barriers to adopting preservation policies at your institution?

Q23: Earlier you indicated that your institution uses the following digital repository or preservation systems: [Qualtrics auto-filled with respondent’s answer from Q2]
What is lacking/missing in the technology being used at your institution?

Q24: Type of institution [Options included: Academic, Archives, Corporate, Government, Museum, Non-profit, Public Library, Other (please describe)]

Q25: Where is your institution located?

Q26: What best describes your role at your institution? [Options included: Administrator, Department or Unit Head, Staff]

Q27: Which of the following best describes your day-to-day work? Select all that apply. [Options included: I do digital preservation, I write policies related to digital preservation, I supervise others who do digital preservation]

Q28: Would you be willing to talk more about your digital preservation practices with us? If so, please provide your name and email address.

Appendix B: Survey Visualization

Appendix C: Interview Questions

Thank you for agreeing to speak with us today. [Introductions of interviewers, including identifying who is the interviewer and who is the note-taker]

This interview is part of "Beyond the Repository," an IMLS grant-funded project exploring the integration of local digital repository services with distributed preservation networks. The goal of these interviews is to explore in-depth the themes of the "Beyond the Repository" survey, including interoperability between distributed and local preservation systems, versioning, and workflows related to those processes. We plan to use information from this interview in our final report, but we will not include anything that could be used to identify you or your institution. This interview should take about an hour, and you can stop at any time. Do you have any questions before we continue?

General information about your digital preservation program

- Please describe the staff involved in digital preservation at your institution, including number of people, roles, and responsibilities.
    - What is your specific role regarding digital preservation?
    - How is your role situated in the larger organization?
    - What type of IT support do you rely on, and how active are they in your preservation practices?
- Tell us more about the history and extent of the digital preservation program at your institution.
    - For how many years have digital preservation activities been undertaken?
    - How much content does your institution have currently preserved? (terabytes)
    - We'd like to get a sense of the maturity level of your organization's digital preservation practices. Is your institution in the beginning, middle or mature in the preservation services offered? Please explain how.
- What systems or services does your organization use for digital preservation?
    - How do they work together (or not?)
    - Why did your institution decide to adopt that system or service?
    - [If applicable] Please describe your institution's homegrown systems more in-depth
        - What functionality were they designed to deliver?
        - Why did you choose to develop software for this purpose?
        - Does it have features that are missing from other systems?

Discussion of distributed preservation systems

- [If applicable] Tell us about the distributed digital preservation system used at your organization.
- [If applicable] Tell us about the preservation workflows used at your organization to deposit into long term preservation systems.

    - If there are different workflows for different systems or types of materials, please discuss for each.

- [If applicable] If you do not have a distributed digital preservation system at your institution and keep offsite copies, describe your current environment.
    - [If applicable] Do you feel your copy distribution is adequate for long-term preservation? Why or why not?

- Tell us about how your organization's local repository and distributed repository (or alternative storage environments, ex. cloud, offsite) interoperate.
    - Is there any shared metadata?
    - Do you re-use local identifiers or have the distributed digital preservation system mint new ones?
    - How do you track what goes offsite?
    - Is there anything that works particularly well about your environment?

- Many survey respondents mentioned that they have created their own workarounds to integrating their local and distributed (or offsite, etc) preservation systems.
    - Have you or your colleagues created workarounds regarding to this specific issue?
    - If so, please describe these workarounds more fully:
        - What tasks are you trying to accomplish that are not provided by the system your organization has chosen?
        - How did your organization solve the issue?
    - If not, please describe other workarounds (if any) that you're created in order to increase interoperability between your preservation system components.

Versioning

- What kinds of versioning practices does your organization have in place?
    - How frequently is versioning happening?
    - What is changing?
    - Are your versioning practices different for local materials, distributed materials, or metadata?
- How do you decide when you version something and when you don't?
- Has your organization ever audited or tried to restore materials you've placed in a distributed system? What happened?

Preservation and curation

- Does your organization have criteria in a policy document to guide the selection of materials for your digital repository? What are they?
- Does your organization have criteria in a policy document to guide which materials are sent to distributed storage and which are stored locally? What are they?
- What types of preservation metadata is used at your organization?
    - Where is preservation metadata stored?

    - How is preservation metadata created and/or updated?

Final thoughts

- Is there anything else you would like to tell us about your digital preservation program?
    - Do you have any additional thoughts on how various tools or services could be more interoperable?
    - Were there any other pain points we did not discuss?

Appendix D: Interview Consent Form

Interview Consent Form
"Beyond the Repository"

This interview is designed to gather information about your organization's digital preservation activities. Your responses will be used in research exploring the integration of local digital repository services with distributed preservation networks as part of the IMLS-funded grant "Beyond the Repository" (IMLS LG-72-16-0135-16).

Procedures: In this one-hour interview, you will be asked questions regarding digital preservation activities at your organization. Questions will address topics such as interoperability between distributed and local preservation systems, versioning, and workflows related to those processes.

Risks: There are no physical or psychological risks involved in this study.

Benefits: There may be no direct benefits to you by your participation in this interview, although your organization may ultimately benefit from any recommendations or solutions that result from this project.

Confidentiality: The interview will be recorded and transcribed. The "Beyond the Repository" team values your candor in these interviews and will remove any information that could be used to identify you or your organization from the project report. If findings from this study are shared more broadly, such as in a conference presentation, any quotations cited from this interview will be stripped of your identity or institutional affiliation to maintain confidentiality.

Subjects' Rights: Your participation in this interview is voluntary and you are free to withdraw at any time, or to refuse to answer any questions. Participation or withdrawal at any time will not have any adverse effect on you.

Contact Persons: Any questions you may have about this study may be directed to Evviva Weinraub, Northwestern University Libraries Associate University Librarian for Digital Strategies, at evviva.weinraub@northwestern.edu or 847-467-6178.

Consent: I have read this form and the purpose of this interview has been explained to me. I have been given the opportunity to ask questions and my questions have been answered to my satisfaction. If I have additional questions, I have been told who to contact. I agree to participate in the research study described above and will receive a copy of this consent form.

Please type your name in the box below to indicate agreement.

_____

Bibliography

Altman, M. Early Results from Auditing Distributed Preservation Networks. (2012, October 23). Retrieved December 20, 2016, from https://drmaltman.wordpress.com/2012/10/23/225/

Bailey, C. W., Jr. (2011). Institutional Repository and ETD Bibliography 2011. Retrieved August 22, 2017, from http://digital-scholarship.org/iretd/iretd.pdf

BitCurator Consortium. (2017). Workflows. Retrieved February 7, 2017, from https://www.bitcuratorconsortium.org/workflows

Caplan, P., Kehoe, W. R., & Pawletko, J. (2010). Towards Interoperable Preservation Repositories: TIPR. International Journal of Digital Curation, 5(1), 34-45.

Coalition for Networked Information. (2017, May). Rethinking Institutional Repository Strategies: Report of a CNI Executive Roundtable Held April 2 & 3, 2017. Retrieved August 22, 2017, from https://www.cni.org/wp-content/uploads/2017/05/CNI-rethinking-irs-exec-rndtbl.report.S17.v1.pdf

CCSDS Secretariat Space Communications and Navigation Office. (2012). Reference Model for an Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-M-2. . Retrieved September 28, 2017 from https://public.ccsds.org/pubs/650x0m2.pdf

Cramer, T., & Kott, K. (2010). Designing and Implementing Second Generation Digital Preservation Services: A Scalable Model for the Stanford Digital Repository. D-Lib Magazine, 16(9/10). Retrieved November 17, 2017, from http://www.dlib.org/dlib/september10/cramer/09cramer.html

Cruse, P., Sandore, B., & National Digital Information Infrastructure and Preservation Program. (2009). The Library of Congress National Digital Information Infrastructure and Preservation Program. Library Trends, 57(3), 301-314. doi: 10.1353/lib.0.0055

Davis, C., & Council of Prairie and Pacific University Libraries. (2016, February 19). Recommendations Report for the COPPUL Digital Preservation Network. Retrieved February 7, 2017, from http://www.coppul.ca/sites/default/files/uploads/RecommendationsReportfortheCOPPULDigitalPreservationNetwork_0.pdf

Eckard, M., Pillen, D., & Shallcross, M. (2017-01-30). Bridging Technologies to Efficiently Arrange and Describe Digital Archives: the Bentley Historical Library's ArchivesSpace-Archivematica-DSpace Workflow Integration Project. Code4Lib Journal, (35). Retrieved February 7, 2017, from http://journal.code4lib.org/articles/12105

Educopia Institute. (2010). A Guide to Distributed Digital Preservation (K. Skinner & M. Schultz, Eds.). Retrieved October 5, 2017, from https://metaarchive.org/wp-content/uploads/2017/03/A_Guide_to_Distributed_Digital_Preservation_0.pdf

Hall, N., & Boock, M. (2017). Environmental Scan of Distributed Digital Preservation Services: A Collective Case Study. Proceedings of the 14th International Conference on Digital Preservation, 85-89. Retrieved November 20, 2017 from https://ipres2017.jp/wp-content/uploads/ver09.pdf

Jones, R. (2006). Institutional Repositories. In K. Garnes, A. Landøy, & A. Repanovici (Eds.), Aspects of the Digital Library. Retrieved February 22, 2017, from http://hdl.handle.net/1956/1829

Jordan, C., McDonald, R., Minor, D., & Kozbial, A. (2008). Cyberinfrastructure Collaboration for Distributed Digital Preservation. 2008 IEEE Fourth International Conference on eScience, 408-409. doi:10.1109/eScience.2008.163

Kunze, J., Littman, J., Madden, L., Summers, E., Boyko, A., & Vargas, B. (2016, January). The BagIt File Packaging Format (V0.97). Retrieved September 12, 2017, from https://tools.ietf.org/html/draft-kunze-bagit-13

LeFurgy, W. B. (2002). Levels of Service for Digital Repositories. D-Lib Magazine, 8(5). Retrieved November 17, 2017 from http://www.dlib.org/dlib/may02/lefurgy/05lefurgy.html

Li, Y., & Banach, M. (2011). Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries. D-Lib Magazine 17(5/6). Retrieved November 17, 2017 from http://www.dlib.org/dlib/may11/yuanli/05yuanli.html

Lynch, C., & Lippincott, J. (2005). Institutional Repository Deployment in the United States as of Early 2005, D-Lib Magazine, 11(9). Retrieved December 20, 2016, from http://http//www.dlib.org/dlib/september05/lynch/09lynch.html

Minor, D., Sutton, D., Kozbial, A., Burek, M., & Smorul, M. (2010). Chronopolis Digital Preservation Network. International Journal of Digital Curation, 5(1), 119-133. Retrieved December 20, 2016, from: http://www.ijdc.net/index.php/ijdc/article/view/150/212

Payette, S., & Lagoze, C. (1998). Flexible and Extensible Digital Object and Repository Architecture (FEDORA). In Nikolaou, C., & Stephanidis, C. (eds) Research and Advanced Technology for Digital Libraries. ECDL 1998. Lecture Notes in Computer Science, vol 1513. doi: 10.1007/3-540-49653-X_4

Purdue University. (2016). Purdue University Workflow. Retrieved February 7, 2017, from https://www.bitcuratorconsortium.org/workflows/purdue-university-workflow

Reich, V.A., & Rosenthal, D. (2009). Distributed Digital Preservation: Private LOCKSS Networks As Business, Social, and Technical Frameworks. Library Trends, 57(3), 461-475. doi: 10.1353/lib.0.0047

Rosenthal, D., & Vargas, D. (2013). Distributed Digital Preservation in the Cloud. International Journal of Digital Curation, 8(1), 107-119. doi: 10.2218/ijdc.v8i1.248

Sites, M. (2013, April 22). The APTrust Story. Retrieved February 7, 2017, from http://aptrust.org/aptrust-admin/resources/the-aptrust-story.pdf

Skinner, K., & Halbert, M. (2009). The MetaArchive Cooperative: A Collaborative Approach to Distributed Digital Preservation. Library Trends, 57(3), 371-392. doi: 10.1353/lib.0.0042

Smith, A. (2006). Distributed Preservation in a National Context. D-Lib Magazine, 12(6). Retrieved December 20, 2016, from http://www.dlib.org/dlib/june06/smith/06smith.html

Tananbaum, G. (2009). Institutional Repositories: The Promises of Yesterday and of Tomorrow. ALCTS 2009 Midwinter Symposium Keynote Address. Retrieved February 22, 2017, from http://www.ala.org/alcts/confevents/upcoming/webinar/irs/040809past

Walters, T., Bishoff, L., Gore, E.B., Jordan, M., & Wilson, T.C. (2009). Distributed Digital Preservation: Technical, Sustainability, and Organizational Developments. California Digital Library. UC Office of the President: California Digital Library. Retrieved December 20, 2016, from: https://escholarship.org/uc/item/38g232wc

Zierau, E., & McGovern, N. Y. (2014). Supporting the Analysis and Audit of Collaborative OAIS's Using an Outer OAIS-Inner OAIS (OO-IO) Model. Proceedings of the 11th International Conference on Digital Preservation, 209-218. Retrieved August 22, 2017, from https://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version_1.pdf