Search D-Lib: Go!

D-Lib® Magazine

The Magazine of Digital Library Research

HOME | ABOUT D-LIB | CURRENT ISSUE | ARCHIVE | INDEXES | CALENDAR | AUTHOR GUIDELINES | SUBSCRIBE | CONTACT D-LIB

D-Lib Magazine

NDSA Storage Report: Reflections on National Digital Stewardship Alliance Member Approaches to Preservation Storage Technologies

Micah Altman
MIT Libraries
micah_altman@alumni.brown.edu

Jefferson Bailey
Metropolitan New York Library Council
jbailey@metro.org

Karen Cariani
WGBH Media Library and Archives
karen_cariani@wgbh.org

Michelle Gallinger, Jane Mandelbaum, Trevor Owens
Library of Congress
{mgal, jman, trow}@loc.gov

Printer-friendly Version

Abstract

The structure and design of digital storage systems is a cornerstone of digital preservation. To better understand ongoing storage practices of organizations committed to digital preservation, the National Digital Stewardship Alliance conducted a survey of member organizations. This article reports on the findings of the survey. The results of the survey provide a frame of reference for organizations to compare their storage system approaches with NDSA member organizations.

Introduction

The National Digital Stewardship Alliance (NDSA) is a network of partners dedicated to ensuring enduring access to digital information. The Alliance's mission is to establish, maintain, and advance the capacity to preserve our nation's digital resources for the benefit of present and future generations.

NDSA membership consists of universities, consortia, professional societies, commercial businesses, professional associations, and government agencies at the federal, state and local level. The Alliance sponsors working groups that enrich digital preservation practice for all. The NDSA Infrastructure Working Group identifies and shares emerging practices around the development and maintenance of tools and systems for the curation, preservation, storage, hosting, migration, and similar activities supporting the long term preservation of digital content. In one effort to achieve that, the NDSA Infrastructure Working Group conducted a

member survey, examining trends in preservation storage. The goal of this survey was to develop a snapshot of storage practices within the organizations of the NDSA. This survey is part of the group's larger effort to explore NDSA members' current approaches to large-scale storage systems for digital stewardship as well as the potential for cloud computing and storage in digital preservation.

The NDSA storage survey was conducted between August 2011 and November 2011. Responses were received from 58 of the 74 members, who stated that they were actively involved in preserving digital content at the time. This represents a 78% response rate. NDSA had a total of 98 members during that period. The respondents represent a diverse cross section of organizations working with preservation storage systems. A copy of the survey can be found in Appendix 1, and a Glossary of Terms used in this report in Appendix 2.

Diverse Partnership, Common Goals

The partners who responded to the questionnaire illustrate the diversity of the National Digital Stewardship Alliance. They include federal and state agencies, public and commercial media organizations, research libraries, and non-profit organizations tasked with the stewardship of digital information. Each partner has different specific stewardship goals (for example, re-use, public access, internal access, legal mandate, etc.). With that noted, all have a goal to not only preserve but make accessible their digital content in order to help their organization fulfill its mission.

Partners are storing a wide range of digital materials. Nearly all reported a significant amount of text and still images. Many are also storing or beginning to store moving images and audio files. For the most part, members described their 2011 collections as between 50-400 TB of digital materials, although one respondent is storing 5 PB. Partners preserving text documents reported much larger object counts than those with other media.

Nearly all respondents are using some sort of commercial spinning disk/server storage for online storage in combination with a digital data tape storage system for offline/nearline storage. Within the NDSA, many members also participate in a range of distributed replication/infrastructure networks and cooperatives (i.e. LOCKSS, MetaArchive, Data-PASS, etc.).[1]

Key Findings

The key findings from the survey were:

- 90% of respondents are distributing copies of at least part of their content geographically.
- 88% of respondents are responsible for their content for an indefinite period of time.
- 80% of respondents use some form of fixity checking for their content.
- 75% of respondents report a strong preference to host and control their own technical infrastructure for preservation storage.
- 69% of respondents are considering, or currently participating in, a distributed storage cooperative or system (ex. LOCKSS alliance, MetaArchive, Data-PASS).
- 64% of respondents are planning to make significant technological changes in their preservation storage architecture in the next three years.
- 51% of respondents are considering or already using a cloud storage provider to keep one or more copies of their content.
- 48% of respondents are considering, or currently contracting out, storage services to be managed by another organization or company.

Note that these percentages varied by organizational role, although in general the subpopulations are too small to support reliable inferences about the differences. In Appendix 3, we provide details on the distribution of key findings by organizational role, and of selected other responses. Note that not all respondents responded to all questions, although question-level non-response was generally quite low. Throughout this article, proportions reported are calculated as a percentage of those responding to the specified question. And to support further analysis of response rates, replication and reanalysis, we have deposited a de-identified open access version of the response data in a public archive.[2]

Trends in the survey responses are grouped in four areas:

- Diversity of Access Requirements
- Distributed and Remote Preservation Storage
- File Fixity and Digital Preservation Storage

- Infrastructure Plans

Diversity of Access Requirements

Diversity is the primary feature of NSDA members' approaches to and requirements for access. Member organizations are providing very different degrees of access to their holdings, managing everything from currently inaccessible dark archives to various modes of offline and online access, as well as support for high performance computing usage. Access ranged from very low to very high availability, and can be described using five categories: *dark archives, offline availability, nearline availability, online availability, and high-performance availability*.

The responses indicated that:

- 59% of the responding members have collections with requirements for instant access to a moderate number of simultaneous users necessitating online availability.
- 40% of the responding members have collections that are kept for eventual availability only. These collections are dark archives or are being kept strictly for disaster recovery.
- 28% of the responding members have collections needing nearline availability, meaning the ability to retrieve content within three hours of a request.
- 24% of the responding members have collections requiring retrieval within two business days of a request; this allows for offline availability.
- 21% of the responding members have collections that require high-performance availability, which includes access to large numbers of simultaneous users or for high-performance computing.

There is also substantial diversity in the access requirements within each organization. Member organizations frequently provide different levels of access for different collections they hold. For example, an organization may need to provide high availability for some collections and low availability for other collections. Among the five categories of access requirements bulleted in the above list:

- 53% reported having a single access level (e.g. online availability) for all the collections they are preserving. That is, just over half of the organizations are providing a single degree of access to all of their materials.
- 31% reported supporting two degrees of access among their collections (e.g. dark archives and online availability depending on the collection).
- 16% reported supporting three or more degrees of access among their collections.

Additionally, many members have different storage systems for preservation and access. A majority of the organizations are providing separate systems for preservation and access. Indeed, 65% of the organizations reported using separate systems, while only 35% reported using the same system for both preservation and access.

Distributed and Remote Preservation Storage

General conversation about "the cloud" in information technology tends to focus on third-party cloud storage providers. Adoption of these cloud storage services remains relatively small. However, when we consider cloud storage alongside several related ways of distributing and using storage as a service, some interesting trends emerge. The answers illuminate both the widespread acceptance of some digital preservation storage practices and the continuing uncertainty regarding others. For example, there is broad acceptance of the importance of geographic redundancy in maintaining preservation copies of content. A majority of members are currently keeping all or some of their preservation copies in multiple geographic locations. This geographic redundancy of digital content signals a success in establishing baseline best practices for preservation storage.

Similarly, participation in distributed or collective preservation systems is gaining in popularity, with half the respondents participating in or planning on joining such a system. Lastly, usage of third-party and cloud-based storage systems is still a disquieting idea. Many members are exploring this option, but functionality challenges, issues of trustworthiness, and uncertainty over sustainability are limiting widespread adoption.

- 76% report keeping data in more than one location for all their content.
- 14% reported keeping a complete copy in multiple geographic locations for some of their content.
- 10% reported that they do not keep their data in multiple geographic locations.

Cooperatives, Contracting Out, and Cloud Storage

Members were asked if they were currently using, planning to use, exploring the possibility of using, or not considering using a distributed storage cooperative, a contracted provider of storage, or other third party cloud storage providers. The chart below reports the members' responses.
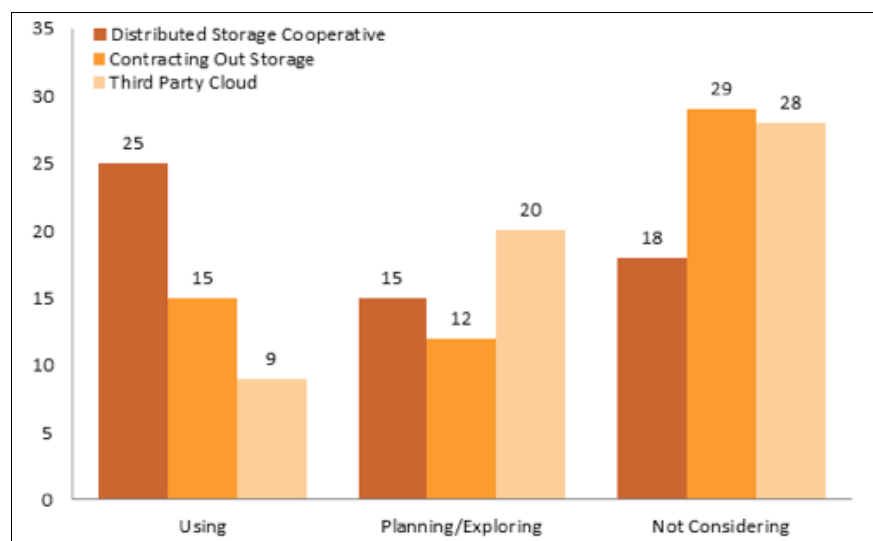

Figure 1: Members' Dispositions to Cooperatives, Contracting, and Cloud Storage

Among the membership there is an implied trend toward participating in distributed storage cooperatives. There is also a substantive interest in cloud storage illustrated by the 20 members currently exploring or planning on incorporating cloud storage systems.

Distributed Storage Cooperatives or Systems

- 43% are using distributed and/or cooperative systems.
- 26% are planning on or actively exploring using these systems.
- 31% of members are not currently considering this storage option.

Of those using or exploring distributed and cooperative systems, 67% are using or exploring some type of LOCKSS system (LOCKSS was used by 81% of respondents reporting they used a distributed/cooperative system). The trustworthiness of distributed digital preservation cooperatives appears to be gaining acceptance.

Contracting Out Storage Services

- 27% are currently contracting out some of their preservation storage to third parties.
- 4% are planning to contract out some of their preservation storage.
- 18% are currently exploring this option.
- 51% of members are not considering contracting out storage services to be managed by a third party.

Third-Party Cloud Storage Service Providers

- 16% of members are using third-party cloud storage service providers for keeping at least one copy of their content.
- 7% are planning on using third-party cloud storage service providers.
- 28% are currently exploring this option.
- 49% are not considering using cloud services for keeping any copies of their content.

Control and the Cloud

The survey revealed the tension between using third-party systems and a preference to host, maintain, and control preservation storage by the organizations themselves. Nearly 50% of respondents are using, planning on using, or considering contractor services or third-party cloud storage. At the same time, 74% of the members agreed or strongly agreed that they had a strong preference for maintaining and controlling their preservation storage systems. The most-cited reasons for this preference were costs, trustworthiness, legal mandate, and security and risk management.

One survey question offers insight into this seeming contradiction. The question asked members to rank the significance of specific preservation storage system features (with 1 being least significant and 7 being most significant). The chart below shows the results. The highlighted cells indicate 10 or more responses. The sum for each function was calculated by multiplying the number of responses by the priority and then adding the totals.

Table 1: Priorities for Functionality in New Storage Systems Sorted by Sum of Priority Scores

Functionality	Priority Scores (1 low, 7 high)							Sum of Priority Scores
	1	2	3	4	5	6	7	
More built-in functions (like fixity checking)	0	3	0	9	14	16	13	**299**
More automated inventory, retrieval and management services)	0	1	5	7	17	10	15	**295**
More Storage	0	4	4	5	12	12	17	**291**
Higher performance processing capacity (ex. indexing on content)	0	3	5	6	21	13	6	**270**
File format migration	2	3	4	10	17	9	9	**261**
More security for the content	1	5	6	15	13	8	8	**258**
Block level access to storage	11	16	5	11	7	4	0	**161**

There is strong demand for features that contractor and third-party cloud services are not yet widely satisfying: primarily built-in fixity checking, automated tasks, and migration services; and secondarily, block-level access. There is not as strong an interest by the preservation community in the block-level access feature as in the larger cloud community. Instead, the NDSA member organizations represent a group looking for a degree of granularity of control over their data that is not widely shared by organizations that do not have a preservation focus. This is also reflected by the higher participation in distributed cooperative systems. Vendor and cloud-based systems are playing a significant role in preservation, but a dearth of functionality and the uncertainties inherent in relinquishing control are likely limiting their widespread use.

Taking this information about desired functionality into account, these results suggest that views of control are being expressed in different ways. While to some, control may mean block-level access to content, this level of control was far and away the least requested feature. In contrast, to the digital preservation community of the NDSA, built-in functions like fixity checking and automated inventory, retrieval and management services express a different sense of control. Built-in functionality provides an organization with preservation information that gives assurance over the integrity of their content. However, it does so by actually reducing the direct control individuals in the organization can exert on digital objects. In this sense, control may be conditionally defined according to specific preservation activity or whether that activity is occurring locally or in "the cloud." Here the survey results open up more questions than they answer about exactly what kinds of control member organizations want to be able to exercise. The combined desire for control over storage coupled with a desire for additional automated functionality suggests that desires for control are not manifesting themselves in strong desires for block-level control.

File Fixity and Digital Preservation Storage

Digital objects pose difficulties to ensuring their ongoing authenticity and stability. Files can become corrupted by use, bits can rot even when unused, and during transfer the parts essential to an object's operability can be lost. At the most basic level, digital preservation requires us to be confident that the objects we are working with are the same as they were prior to our interaction with them.

To deal with this problem, those in the digital preservation field often talk about the fixity of digital objects. Fixity, in this sense, is the property of being constant, steady, and stable. Content stewards can check their digital objects to make sure that they maintain these qualities. Fixity checking is the process of verifying that a digital object has not been altered or corrupted. In practice, this is most often accomplished by computing and comparing cryptographic hashes (these are sometimes loosely referred to as "checksums").

NDSA Members' Approaches to Fixity Checking

One key theme that emerged from the survey was the prevalence of fixity checking as a performance requirement and the challenges imposed on storage systems by this activity.

Eighty-eight percent of the responding members are doing some form of fixity checking on content they are preserving. This widespread use of fixity checking illustrates the recognition that validation of the integrity and consistency of the objects we are preserving is a critical component in digital preservation workflows.

With that said, NDSA members are taking distinctly different approaches to checking the fixity of their content. The differences are most likely due to a variety of complicated issues including the scalability of fixity-checking software, network limitations and data transfer costs, transaction volume and access requirements, and other contextual factors around the availability and management of specific sets of content. Amongst survey respondents, fixity checking occurs as follows, with some members maintaining multiple practices:

- 82% of the organizations report that they are doing some form of fixity checking on content they are preserving.
- 57% of the organizations are doing checks before and after transactions such as ingest.
- 34% of the organizations are doing checks on some reoccurring fixed schedule.
- 32% of the organizations are randomly sampling their content to check fixity.
- 18% of the organizations use tamper-resistant fixity check mechanisms.[3]
- 17% of the organizations store fixity information in an independent system.

Most respondents reported using multiple practices.

While fixity checking itself is widespread, NDSA members also take various approaches to scheduling these checks. Some are randomly sampled and others use a fixed schedule for checking. Twenty-four of the responding organizations use a fixed schedule for at least part of their content.

- 46% check fixity of content on at least a monthly basis.
- 21% check fixity of content on at least a quarterly basis.
- 29% check fixity of content on an annual basis.
- 4% check fixity of content on a tri-annual basis.

The Future of Fixity

NDSA Infrastructure working group members have frequently noted that the state of the art in fixity checking involves distributed fixity checking and frequent, robust repair to intentional or unintentional corruption. This is done by replacing corrupted data with the distributed, replicated, and verified data held at "mirroring" partner repositories in multi-institutional, collaborative distributed networks. The consortia groups MetaArchive and Data-PASS use LOCKSS for this kind of distributed fixity checking and repair. These consortia and a number of others are also using or testing the SafeArchive tool, developed by NDSA members, which provides an automated collection-fixity and replication-policy auditing on top of distributed storage networks such as LOCKSS.[4] As well, some individual institutions use a self-maintained distributed repository system that allows them to replace damaged content with a verified, uncorrupted copy; or are investigating services such as DuraCloud[5] that provide at least some fixity checking services.

As previously mentioned, one of the key interests of this NDSA working group was the potential role for cloud storage systems in digital preservation storage architectures. For those using cloud storage systems, complying with fixity requirements can prove problematic. As David Rosenthal has suggested in 2011[6], cloud services at the time were not able to prove that they are not simply replaying fixity information created and stored at the time of deposit. Rosenthal highlighted the need for cloud services to provide a tool or service to verify that the systems hold the content rather than simply caching the fixity metadata. Without that kind of assurance, it can be prohibitively expensive to run any kind of frequent fixity checks on content in various cloud storage platforms.

Built-in functionality like automated fixity checking and repair was highlighted as the most desired feature in future preservation storage systems. This desire, along with the challenges of system-type dependencies and diversity of uniform current practices in fixity checking procedures, show the complex interplay between access, performance, preservation requirements, storage infrastructure, and institutional resources. As practices such as fixity checking become ubiquitous and new options like distributed storage gain further acceptance, the hardware underpinning these requirements will be called upon to meet new demands. Our hope is that preservation stewards navigating these decisions will benefit from the knowledge and experience of other NDSA members as they encounter similar complexities and devise new solutions.

Infrastructure Plans

There are a number of survey questions which did not fit thematically in early sections but will be of interest to both content users and service providers — specifically storage media currently being used by survey respondents, how many preservation copies of digital assets institutions are keeping, and the number of members that have documented requirements for storage systems.

Number of Copies, Storage Media, and Documented Requirements

The chart below shows the number or preservation copies institutions are keeping, with 45% keeping three or more copies of their digital assets. (See Appendix 2 for a distribution by organization type.)
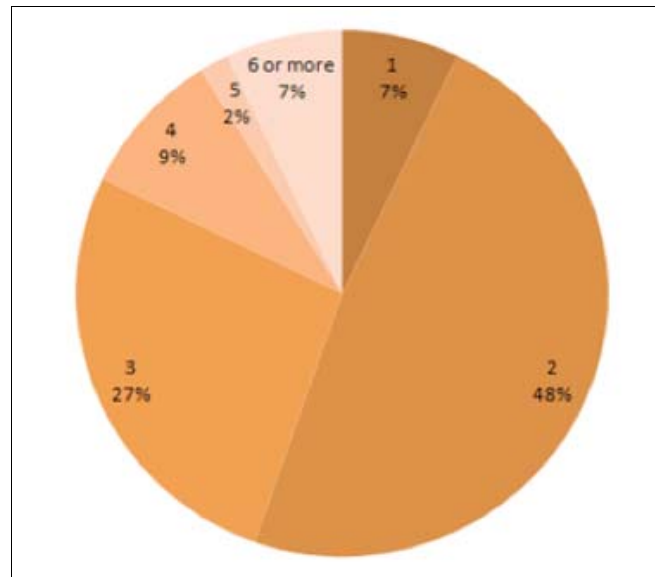

Figure 2: Number of preservation copies of digital assets survey participants are keeping

Figure 3 shows the media being used by members for preservation storage. Some members use multiple kinds of media.
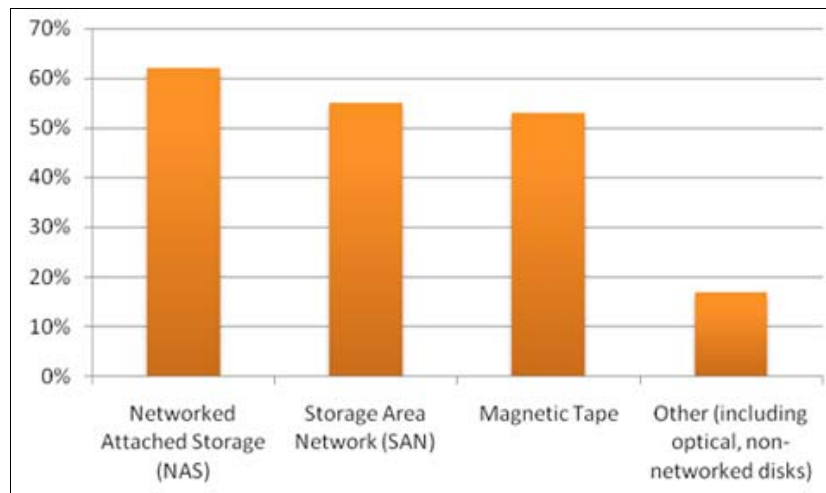

Figure 3: Percentages of Types of Storage Media NDSA Members Use for Preservation Storage

The question "Does your organization have specific documented requirements for your storage systems" elicited a wide range of responses to the different types of requirements. Forty-nine of the 58 organizations that responded to the survey reported currently having some form of requirements, or planning to develop requirements in the next year.

Within this subset of respondents, the specific requirements varied:

○ 43% have documented functional requirements.

- 37% have documented security requirements.
- 35% have documented general performance requirements.
- 29% plan to develop requirements within one year.
- 18% have other documented requirements.
- 16% have documented performance requirements for ingest.
- 12% have documented performance requirements for migration to new technology or other one-time intensive operations.

For the 18% claiming "other documented requirements," the additional document requirements were most often client-specific or content-specific.

Storage Usage and Expectations

One fundamental consideration when planning digital preservation infrastructure needs is the amount of storage space required. The survey queried participants both on the amount of storage space they were currently using for all copies of their digital content and the amount they expect to need three years from now.

Table 2: Storage Use and Expectations

Storage Space Amount	Current Storage For All Copies	Requirement Anticipated in 3 Years For All Copies
Under 10 TB	18	13
10-99 TB	19	13
100 to 999 TB	14	16
1000+ TB (1+ PB)	5	9

Charting out these numbers shows the expected growth of storage needs in the next three years, especially in the upper ranges of storage amount. The chart shows many of the member organizations moving out of the less than ten terabytes category and moving into the bigger brackets. Notably, the 1000+ TB (1+ PB) category is likely to see the largest increase, almost doubling from 5 members to 9.

When averaged out between the two questions, the disparity in the amount of storage used in 2011 and expected to be needed in 2014 becomes even more apparent. The 2011 usage averaged out to 492 TB per institution whereas anticipated need in three years more than doubled, averaging out to 1107 TB per institution.

Predicting Future Storage Needs

A number of the survey questions asked members to estimate other aspects of digital preservation storage needs three years in the future. While cost modeling for digital preservation has been getting increased research scrutiny lately,[7] the Storage Survey polled members on issues of strategic planning and administration of infrastructure including expectations on technology changes, available resources, organizational plans, and audit and certification as a trustworthy repository.

The speed of technological change and its impact on digital preservation is nowhere more evident than in the fact that 64% of respondents agree or strongly agree that their organization plans to make significant changes in technologies in their preservation storage architecture within the next three years. At the same time, survey participants remained confident of their ability to meet these challenges, with 83% agreeing or strongly agreeing their institution will have adequate resources to meet projected preservation storage requirements over the next three years.

Table 3: Future Storage Needs

In the next three years my organization plans to…	Agree	Neutral	Disagree
Make significant changes in preservation storage technologies	37 (64%)	10 (17%)	11 (19%)
Will have adequate resources to meet storage requirements	48 (83%)	7 (12%)	3 (5%)
Has a plan to meet our preservation storage requirements	45 (79%)	8 (14%)	4 (7%)

Plans to meet trustworthy digital repository requirements	32 (57%)	19 (34%)	5 (9%)

As is evident in the table, the statistics on adequate resources expectations and proper organizational planning are very similar. The positivity reflected in these numbers is a good sign for the future of digital preservation. Another positive result revolved around expectations for meeting the requirements for the recently approved ISO Standard 16363, or related Trustworthy Repositories Audit & Certification.[8] The fact that 60% of the survey respondents plan on complying with the rigorous TRAC standards within three years signals an increased acknowledgement of the importance of these requirements in certifying digital preservation repository standards.

Conclusions

Survey respondents did not shrink from the challenges of meeting the requirements, current and future, of digital preservation storage. The survey revealed an inherent optimism in addressing future digital preservation storage infrastructure issues even as anticipated storage needs rise dramatically and technology changes often. The results also revealed the complexity of digital preservation storage planning, especially given the large number of preservation copies being maintained, diversity of media used, and access requirements documented. The survey results communicated that organizations committed to the long-term preservation of digital materials share concerns and needs across industries. These needs are similar but not identical to those of organizations whose mission does not include providing long-term access to data.

Next Steps

The NDSA plans to reissue the survey on a periodic basis, to track the trends and requirements of the membership and provide useful information to others in the community as well as service providers. Feedback from the members will be incorporated to enhance the survey questions. For example, one question not included in the survey, but that will be considered for future storage surveys, is inquiring if institutions are planning on maintaining the same number of file copies into the future or whether redundancy policies are flexible in response to infrastructure limitations or forecasting. As the size of each digital item increases, and as size-intensive formats like audio and video become a larger percentage of preserved collections, keeping multiple copies will have an increased impact on storage capacity needs. Other potential areas of investigation for a future survey could revolve around the roles that formats, compression, and means of access play in determining storage infrastructure.

As institutions plan for their future storage needs, the knowledge sharing and collaboration activities of the NDSA will offer guidance as they make digital preservation infrastructure decisions.

Acknowledgement

The authors would like to thank the members of the NDSA Infrastructure Working Group.

References

[1] For descriptions of these networks see:

- Reich, Vicky & Rosenthal, David S.H., "LOCKSS: A Permanent Web Publishing and Access System", *D-Lib Magazine*, vol. 7, no. 6, June 2001, http://dx.doi.org/10.1045/june2001-reich.
- T. Walters & K. Skinner (Eds.), New Roles for New Times: Digital Curation for Preservation (pp. 51—53), Association of Research Libraries.
- Altman, M. (2009), Digital Preservation through Archival Collaboration: The Data Preservation Alliance for the Social Sciences, *American Archivist*, 72(72, no. 1 (2009)), 170—184.

[2] Altman, Micah; Bailey, Jefferson; Cariani, Karen; Gallinger, Michelle; Owens, Trevor (2012), "Data for NDSA Storage Report 2011", http://hdl.handle.net/1902.1/19768 V1.

[3] This refers to local direct use of tamper-resistant mechanisms. Indirect use is higher — as reported above, 25 organizations (43%) participate in a distributed storage cooperative, and over 80% of these cooperatives (34% of the total number of respondents) use tamper resistant fixity-check mechanisms.

[4] Altman, M., & Crabtree, J. (2011). Using the SafeArchive System: TRAC-Based Auditing of LOCKSS. *Archiving 2011* (pp. 165—170). Society for Imaging Science and Technology.

[5] See DuraCloud Health Checkup.

[6] Rosenthal, David S. H., LOCKSS In The Cloud, presented at "Make It Work: Improvisations on the Stewardship of Digital Information", Joint NDSA NDIIPP partners meeting, July 19-21 2011. Washington, DC.

[7] Three recent recourses for cost modeling digital preservation of interest are:

- Financial Models and Calculation Mechanisms, Hans Westerhof, Jeff Ubois, Marius Snyders, Presto Centre, 2011.
- CDL Cost Modeling, Stephen Abrams, 2012.
- Talk at PDA2012, David Rosenthal, DSHR's Blog, 2012.

[8] Trustworthy Repositories Audit & Certification (TRAC) Criteria and Checklist, Center for Research Libraries & Online Computer Library Center, 2007.

Appendix 1

The NDSA storage survey was conducted between August 2011 and November 2011. Responses were received from 58 of the 74 members who stated that they were actively involved in preserving digital content at the time. There were a total of 98 members of NDSA during that period.

System Survey

The Infrastructure working group of the National Digital Stewardship Alliance is working to better understand how member organizations are approaching storage for their preservation systems. As part of this effort, we are asking each NDSA member institution to respond to this 22-question survey. For institutions where preservation and access are coupled at the storage level, the following questions should be answered for the entire system. For institutions that have separate archival storage, the questions should be answered for the archival storage only.

---

1. My organization's storage system uses the following storage media for preservation storage. (Check all that apply.)

___ Spinning disk — Locally or network attached storage (NAS)

___ Spinning Disk — Storage Area Network (SAN)

___ Magnetic tape

___ Other (specify)

---

2. In general, how many preservation copies of the digital assets are you keeping?

1　2　3　4　5　< 6

---

3. Approximately, how many terabytes of storage space do you require for all copies of your content that you manage?

---

4. Approximately, how many terabytes of storage space do you anticipate needing for all copies of your content that you manage in three years?

---

5. Is your organization keeping copies of digital assets in geographically distinct places to protect from regional geographic disasters? (Check all that apply.)

___ Yes, we manage our own copies in one or more geographically distinct offsite locations

___ Yes, we keep additional copies of our materials in a distributed collaborative partnership

___ Yes, we keep one or more additional copies of our materials managed by another institution or commercial provider

___ In some cases, decided on collection basis

___ No, we would like to but we do not have the resources

___ No, we do not and this is not something we are pursuing

---

6. When does your organization check the fixity of the content you are preserving?

___ We do fixity checks before and after transactions like ingest

___ We do fixity checks on all content we are preserving at fixed intervals

___ We randomly sample content and check for fixity

___ We store fixity information in an independent system

___ We use a tamper-resistant fixity check mechanism (E.g. LOCKSS, ACE)

___ We do not do fixity checks on our content

If your organization performs fixity checks on content you are preserving at fixed intervals, how frequently (in months) do you perform those checks? (i.e., if you perform them monthly, enter 1; if every nine months, enter 9; if annually, enter 12). _____

---

7. Does your organization have specific performance requirements for your storage system or systems? (Check any and all that apply.)

___ We have documented general performance requirements

___ We have documented performance requirements for ingest

___ We have documented performance requirements for migration to new technology or other one-time intensive operations.

___ We have documented functional requirements

___ We have documented security requirements

___ We plan to develop requirements within one year

___ We have other documented requirements (Specify)

---

8. What are your requirements for access to the content you store? (If you have different requirements for different collections please check each option that applies to one of your collections.)

___ Eventual availability only (dark archive/disaster recovery)

___ Off-line availability (e.g. able to retrieve on request w/in 2 business days)

___ Near-line availability (e.g. able to retrieve on request w/in 3 hours)

___ On-line availability (e.g. instant online access for "moderate" number of simultaneous users)

___ High-performance availability (access to large number of simultaneous users/or for HPC)

---

9. Does your organization use separate storage systems for access-only and preservation-only services?

___ Yes

___ No

10. Which services does your organization currently provide for files in your preservation storage? (Check all that apply.)

   ___ Secure storage with backup and recovery procedures in place

   ___ Periodic fixity checking

   ___ Version control

   ___ Format normalization, format migration, or platform emulation

11. Do you provide different services for different "collections" under preservation storage?

   ___ Yes

   ___ No

12. If you do provide different services for different collections please describe them below.

13. How significant are each of the following general features of preservation systems for meeting your organizations objectives? (1 being insignificant, 5 being most significant.)

   ___ More storage

   ___ Block level access to storage (not just file level)

   ___ Higher performance processing capacity (to do processing like indexing on content)

   ___ More built-in functions (like fixity checking)

   ___ More automated inventory, retrieval and management services

   ___ More security for the content

   ___ File format migration

14. My organization has a plan to meet our preservation storage requirements over the next three years.

   ___ Strongly disagree

   ___ Disagree

   ___ Neutral

   ___ Agree

   ___ Strongly agree

   ___ Not Applicable

15. In general, how long (in years) is your organization responsible for preserving content? (Enter 999 if your organization has explicit or implicit indefinite responsibility.)

16. I expect my organization will have adequate resources to meet projected preservation storage requirements over the next three years.

   ___ Strongly disagree

___ Disagree

___ Neutral

___ Agree

___ Strongly agree

___ Not Applicable

---

17. My organization plans to make significant changes in technologies in its preservation storage architecture within the next three years.

___ Strongly disagree

___ Disagree

___ Neutral

___ Agree

___ Strongly agree

___ Not Applicable

---

18. My organization intends to meet requirements for a trustworthy digital repository according to TRAC or the planned ISO standard 16363 within the next three years.

___ Strongly disagree

___ Disagree

___ Neutral

___ Agree

___ Strongly agree

___ Not Applicable

---

19. Is your organization participating in a distributed storage cooperative or system (e.g. LOCKSSAlliance, MetaArchive, Data-PASS)?

___ Yes, my organization currently participates in distributed storage cooperative or system.

___ No, but my organization is planning to participate in a distributed storage cooperative or system.

___ No, but my organization is currently exploring participating in a distributed storage cooperative or system.

___ No, my organization is not considering participating in a distributed storage cooperative or system.

___ No, and my organization is uninterested in participating in a distributed storage cooperative or system.

If you are using, considering or exploring participating in a distributed storage cooperative please list specific cooperative you are participating in, considering or exploring.

---

20. Is your organization contracting out storage services to be managed by another organization or company?

___ Yes, my organization currently contracts out storage services which are managed by another organization.

___ No, but my organization is planning to contract out storage services which are managed by another organization.

___ No, but my organization is currently exploring contracting out storage services which are managed by another organization.

___ No, my organization is not considering contracting out storage services which are managed by another organization.

___ No, and my organization is uninterested in considering contracting out storage services which are managed by another organization.

If your organization is considering, exploring or currently contracting out storage services to be managed by another organization or company please list specific services you are using, considering or exploring.

---

21. Is your organization using third-party cloud storage service providers (e.g. Amazon, Rackspace, Azure, DuraCloud, etc.) for keeping one or more copies of its content?

___ Yes, my organization currently using third-party cloud storage service providers for keeping one or more copies of its content.

___ No, but my organization is planning to use third-party cloud storage service providers for keeping one or more copies of its content.

___ No, but my organization is currently exploring using third-party cloud storage service providers for keeping one or more copies of its content.

___ No, my organization is not considering using third-party cloud storage service providers for keeping one or more copies of its content.

___ No, and my organization is uninterested in using third-party cloud storage service providers for keeping one or more copies of its content.

If you are using, considering or exploring third-party cloud storage service providers (E.g. Amazon, Rackspace, Azure, DuraCloud, etc) for keeping one or more copies of its content please list specific services you are using, considering or exploring.

---

22. My organization has a strong preference to host, maintain, and control its own technical infrastructure for preservation storage.

___ Strongly disagree

___ Neutral

___ Agree

___ Strongly agree

___ Not Applicable

If your organization does have a strong preference to host and control its own technical infrastructure for preservation storage why does it have this preference?

Appendix 2

Glossary of Terms

This glossary lists how terms are used in this document.

Access Storage: Storage designed to contain and serve content to users through common protocols such as the web. Often, this is assumed to be available on a public website (or one accessible to a large group of users such as all students and faculty of a university).

Block-level Access: Reading and writing to disks at the physical level. Only system engineers use block-level access to specify or identify exactly where data are stored, generally for performance reasons.

Dark Archive: Storage designed to be inaccessible (except for authorized storage system managers).

Fixity: The property of being constant, steady, and stable.

Fixity checking: The process of verifying that a digital object has not been altered or corrupted

High-performance availability: It includes access to large numbers of simultaneous users or for high performance computing.

Nearline Storage: Storage generally designed to provide retrieval performance between online and offline storage. Typically, nearline storage is designed in a way that file retrieval is not instantaneous but is available to the user in the same session.

Offline Storage: Storage recorded on detachable media, not under the control of a processing unit (such as a computer).

Online Storage: Storage attached under the control of a processing unit (such as a computer) designed to make data accessible close to instantaneously.

Preservation Storage: Storage designed to contain and manage digital content for long-term use.

Appendix 3

Distribution of Key Responses by Organizational Role

		archive	library	museum	other	service provider
Geographic Replication						
No	%	21.43%	8.00%	25.00%	0.00%	0.00%
	N	3	2	1	0	0
Yes	%	78.57%	92.00%	75.00%	100.00%	100.00%
	N	11	23	3	6	8
Keep Indefinitely						
No	%	0.00%	8.70%	0.00%	33.33%	37.50%
	N	0	2	0	2	3
Yes	%	100.00%	91.30%	100.00%	66.67%	62.50%
	N	14	21	4	4	4
Strong Control						
No	%	14.29%	24.00%	50.00%	33.33%	28.57%
	N	2	6	2	2	2
Yes	%	85.71%	76.00%	50.00%	66.67%	71.43%
	N	12	19	2	4	5
Change Soon						
No	%	21.43%	28.00%	50.00%	66.67%	50.00%
	N	3	7	2	4	4

Yes	%	78.57%	72.00%	50.00%	33.33%	50.50%
	N	11	18	2	2	4

Collaborative Storage

No	%	28.57%	12.00%	25.00%	66.67%	62.50%
	N	4	3	1	4	5
Yes	%	28.57%	24.00%	50.00%	16.67%	25.00%
	N	4	6	2	1	2
Considering	%	42.86%	64.00%	25.00%	16.67%	12.50%
	N	6	16	1	1	1

Cloud Storage

No	%	53.85%	52.00%	50.00%	50.00%	25.00%
	N	7	13	2	3	2
Yes	%	30.77%	44.00%	25.00%	33.33%	25.00%
	N	4	11	1	2	2
Considering	%	15.38%	4.00%	25.00%	16.67%	50.00%
	N	2	1	1	1	4

Third Party Storage

No	%	50.00%	52.00%	50.00%	83.33%	37.50%
	N	6	13	2	5	3
Yes	%	25.00%	24.00%	25.00%	16.67%	12.50%
	N	3	6	1	1	1
Considering	%	25.00%	24.00%	25.00%	0.00%	50.00%
	N	3	6	1	0	4

Distribution of Number of Copies by Organizational Role

The height of each colored bar represents the number of respondents from each organizational role that indicated they kept that specific number of copies. The width of the bars represents the proportion of each group of total respondents.

About the Authors

Micah Altman is Director of Research and Head/Scientist, Program on Information Science for the MIT Libraries, at the Massachusetts Institute of Technology. Dr. Altman is also a Non-Resident Senior Fellow at The Brookings Institution. Prior to arriving at MIT, he served at Harvard University for fifteen years as the Associate Director of the Harvard-MIT Data Center, Archival Director of the Henry A. Murray Archive, and Senior Research Scientist in the Institute for Quantitative Social Sciences. Dr. Altman conducts research in social science, information science and research methods — focusing on the intersections of information, technology, privacy, and politics; and on the dissemination, preservation, reliability and governance of scientific knowledge.

Jefferson Bailey is Strategic Initiatives Manager at Metropolitan New York Library Council. He previously worked in the Office of Strategic Initiatives at the Library of Congress in the National Digital Information Infrastructure and Preservation Program (NDIIPP) and Digital Preservation Outreach and Education (DPOE) program. He has managed digital projects at Brooklyn Public Library and the Frick Art Reference Library and has done archival work at NARA and NASA.
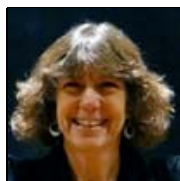
Karen Cariani is the Director of the WGBH Media Library and Archives. Karen has worked at WGBH since 1984 in television production and archival-related roles. She has 20 plus years of production and project management experience, having worked on numerous award-winning historical documentaries including MacArthur, Rock and Roll, The Kennedys, Nixon, and War and Peace in the Nuclear Age. She also worked with the WNET, PBS, NYU and WGBH Preserving Public Television partnership as part of the Library of Congress National Digital Information Infrastructure Preservation Project. She served two terms (2001-2005) on the Board of Directors of the Association of Moving Image Archivists (AMIA). She was co-chair of the AMIA Local Television Task Force, and Project Director of the guidebook "Local Television: A Guide To Saving Our Heritage," funded by the National Historical Publications and Records Commission. She is currently co-chair of the LOC National Digital Stewardship Alliance Infrastructure Working Group.

Michelle Gallinger is Digital Programs Coordinator for the National Digital Information Infrastructure and Preservation Program at the Library of Congress. Gallinger works to develop the digital preservation community, including the planning and execution of various international Aligning National Approaches to Digital Preservation activities. Gallinger develops policies and guidelines for digital preservation practices, life cycle management of digital materials, and stakeholder engagement at the Library of Congress. She also provides strategic planning for the National Digital Information Infrastructure and Preservation Program, a collaborative project that supports a network of partners exploring the capture, preservation and provision of access to a rich variety of digital information. Gallinger developed the initial strategy for, and supported the creation, definition, and launch, of the National Digital Stewardship Alliance in 2010 and is currently the NDSA facilitator. Before joining the Library of Congress, Gallinger developed the Colonial Williamsburg Rockefeller Library digitization and digital stewardship practices and worked at the University of Virginia E-Text Center.

Jane Mandelbaum is currently Manager of Special Projects in the Office of the Director for Information Technology Services at the Library of Congress (LC). She is currently leading and guiding enterprise-wide projects and architecture initiatives for large-scale high-performance digital storage and archiving. She previously served as IT implementation and operations manager for a number of large IT systems at LC, and led a team to establish and operate the Library's end-user computing environment.

Trevor Owens is a Digital Archivist with the National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Office of Strategic Initiatives at the Library of Congress. At the Library of Congress, he works on the open source Viewshare cultural heritage collection visualization tool, as a member of the communications team, and as the co-chair for the National Digital Stewardship Alliance's Infrastructure Working Group. Before joining the Library of Congress he was the community lead for the Zotero project at the Center for History and New Media and before that managed outreach for the Games, Learning, and Society Conference.