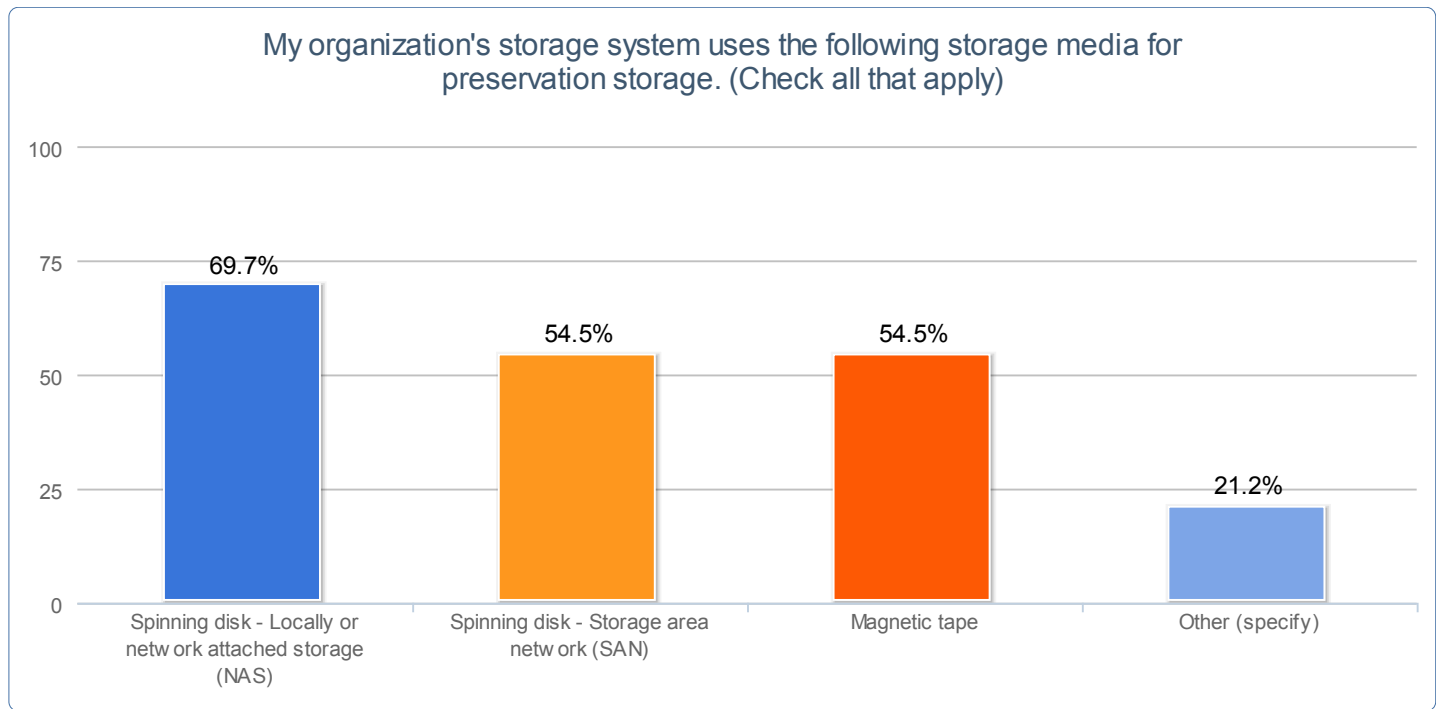# Summary Report - Sep 13, 2011

Survey: NDSA Large Scale Storage System Survey
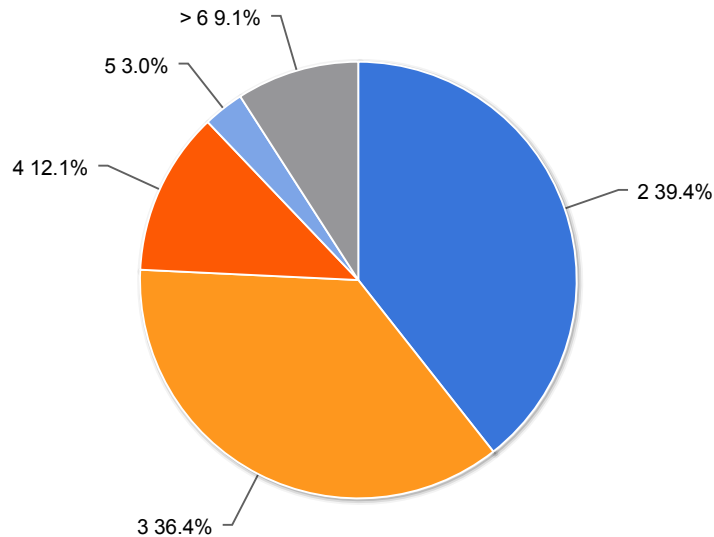
My organization's storage system uses the following storage media for preservation storage. (Check all that apply)

100

75   69.7%

54.5%   54.5%

50

25   21.2%

0
| Spinning disk - Locally or network attached storage (NAS) | Spinning disk - Storage area network (SAN) | Magnetic tape | Other (specify) |

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

Value	Count	Percent %
Spinning disk - Locally or network attached storage (NAS)	23	69.7%
Spinning disk - Storage area network (SAN)	18	54.5%
Magnetic tape	18	54.5%
Other (specify)	7	21.2%

Statistics	
Total Responses	33

Open-Text Response Breakdown for "Other (specify)"	Count
Amazon (S3) & Rackspace (Cloudfiles)	1
CD, DVD, Blu-Ray	1
Cloud storage	1
External hard drives & cloud storage	1
Long-term archival optical disks	1
Looking into buying Millenniata disks	1
remote copy of disk array and tape	1

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

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

> 6 9.1%		
5 3.0%		
4 12.1%		2 39.4%
3 36.4%		

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

Value	Count	Percent %
2	13	39.4%
3	12	36.4%
4	4	12.1%
5	1	3%
> 6	3	9.1%

Statistics	
Total Responses	33
Sum	83.0
Average	2.8
StdDev	0.80
Max	5.0

If you keep different numbers of copies for different assets, please describe.

Count	Response
1	All assets have two copies
1	Depends on library collection policies
1	If an agency keeps one or more copies the archives does not need to have three copies
1	No, in fact we recently made a decision to treat all content exactly the same.
1	Policies are being developed for types of assets
1	Some legacy content--Only because retrospective reingest has yet to happen
1	We keep 3 copies of the preservation copies; and 4 copies of use copies (for faster access)
1	varies by organization unit/pres. system
1	videotape + data file on LTO
1	yes depending on usage and size
1	We keep a third copy on disk for some data simply because it needs to be more immediately available.

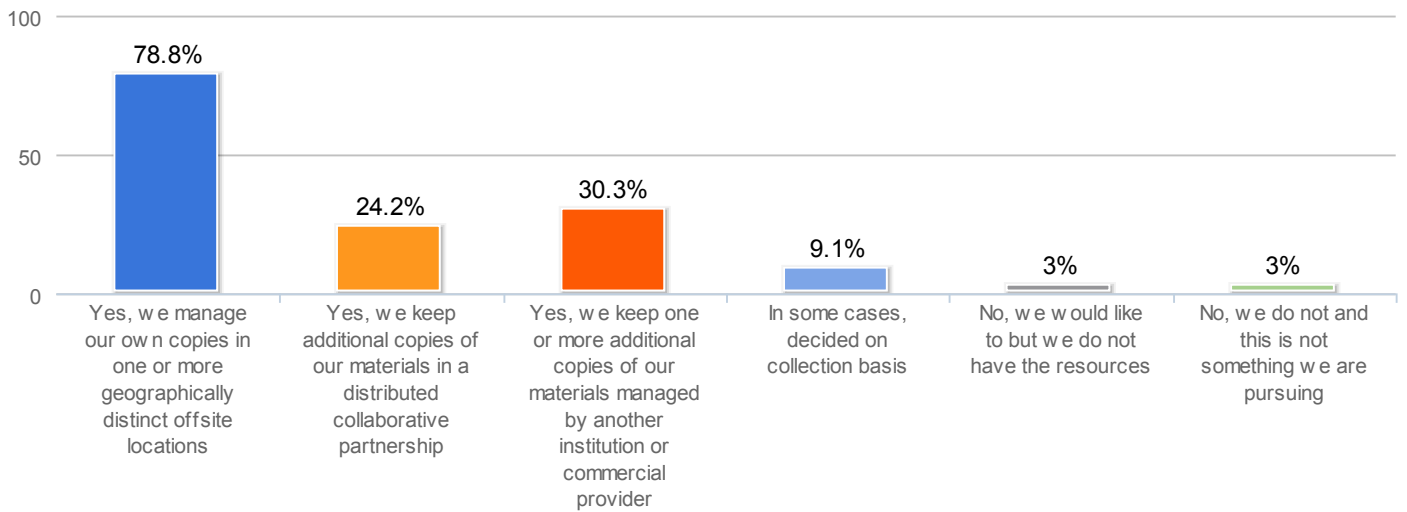|	
1	One local copy, one on server for viewing by users, one extra on server in designated spot inaccessible from web requests.
1	We keep a minimum of 2 replicas of all resources. In general we also keep a 3rd replica with an independent 3rd party and occassionally keep both offline and additional online replicas of some resources in separate repository systems for specific projects to facilitate analysis and access vs preservation.

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

Count	Response
1	1
2	10
1	1000
2	120
1	15
1	150
1	18576
1	2
2	20
1	200
1	250
2	3
3	300
1	4
2	40
1	400
1	426
1	48
1	50
1	60
1	7
1	78
1	80
1	800

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

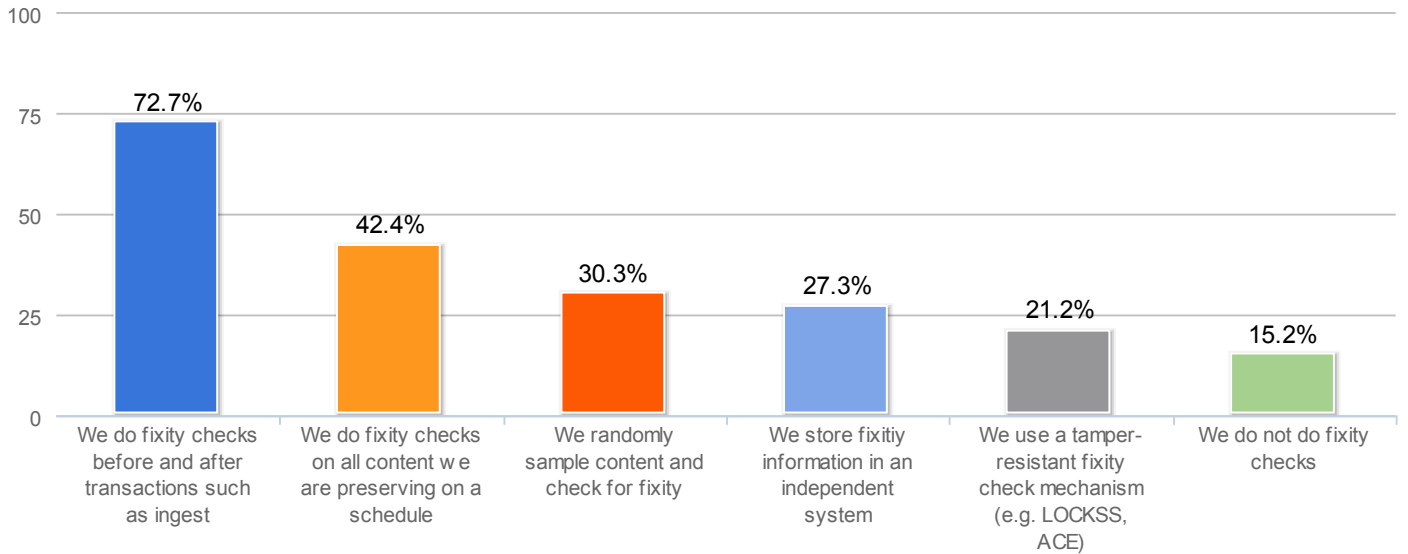| Count	Response
2	1
2	100
1	11
2	150
1	2000
1	31000
1	360
1	3600
2	40
1	400
1	4200
1	45
1	5
3	500
1	5000
3	6
1	60
1	600
1	720
1	732
2	80
1	90

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

Value	Count	Percent %
Yes, we manage our own copies in one or more geographically distinct offsite locations	26	78.8%
Yes, we keep additional copies of our materials in a distributed collaborative partnership	8	24.2%
Yes, we keep one or more additional copies of our materials managed by another institution or commercial provider	10	30.3%
In some cases, decided on collection basis	3	9.1%
No, we would like to but we do not have the resources	1	3%
No, we do not and this is not something we are pursuing	1	3%

Statistics	
Total Responses	33

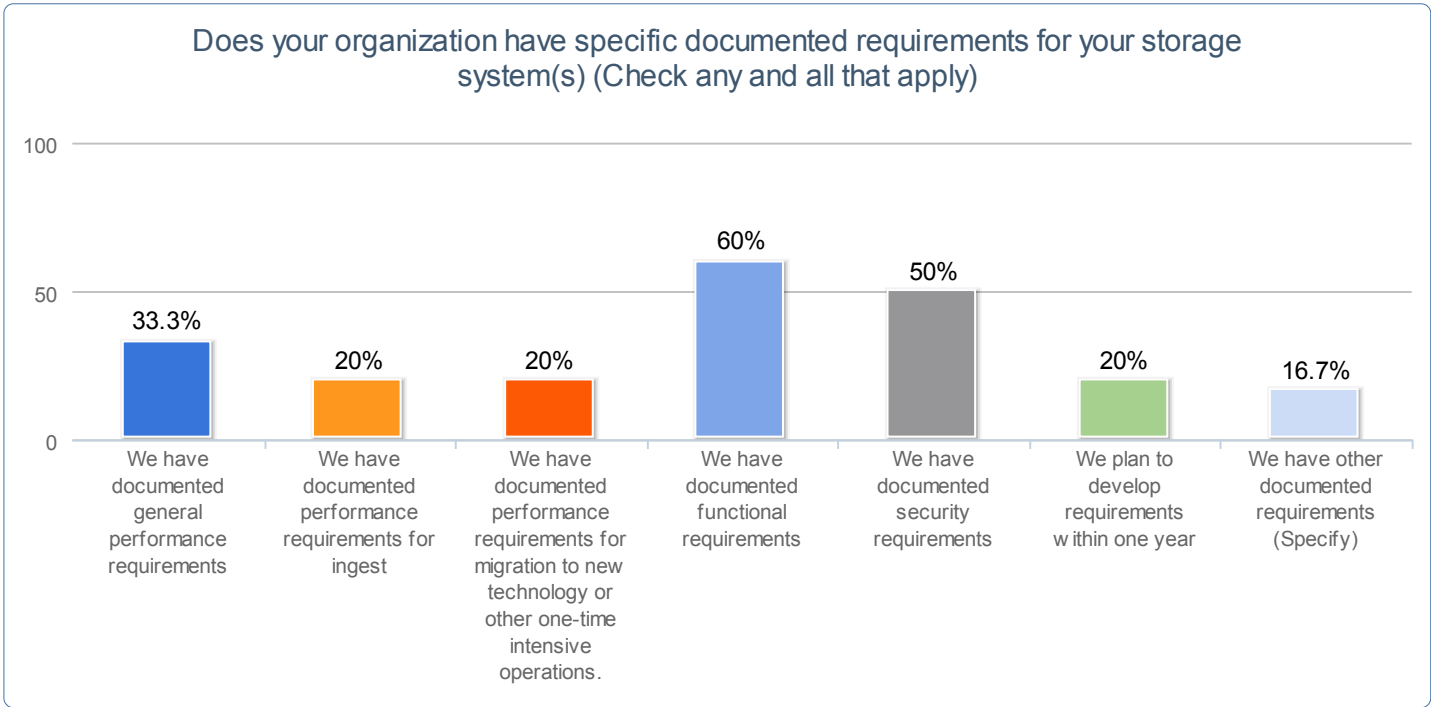When does your organization check the fixity of the content you are preserving?(Check any and all that apply)

When does your organization check the fixity of the content you are preserving? (Check any and all that apply)

Value	Count	Percent %
We do fixity checks before and after transactions such as ingest	24	72.7%
We do fixity checks on all content we are preserving on a schedule	14	42.4%
We randomly sample content and check for fixity	10	30.3%
We store fixitiy information in an independent system	9	27.3%
We use a tamper-resistant fixity check mechanism (e.g. LOCKSS, ACE)	7	21.2%
We do not do fixity checks	5	15.2%

Statistics	
Total Responses	33

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

Count	Response
1	.25
1	0.5
8	1
1	112
2	12
2	2

Does your organization have specific documented requirements for your storage system(s) (Check any and all that apply)
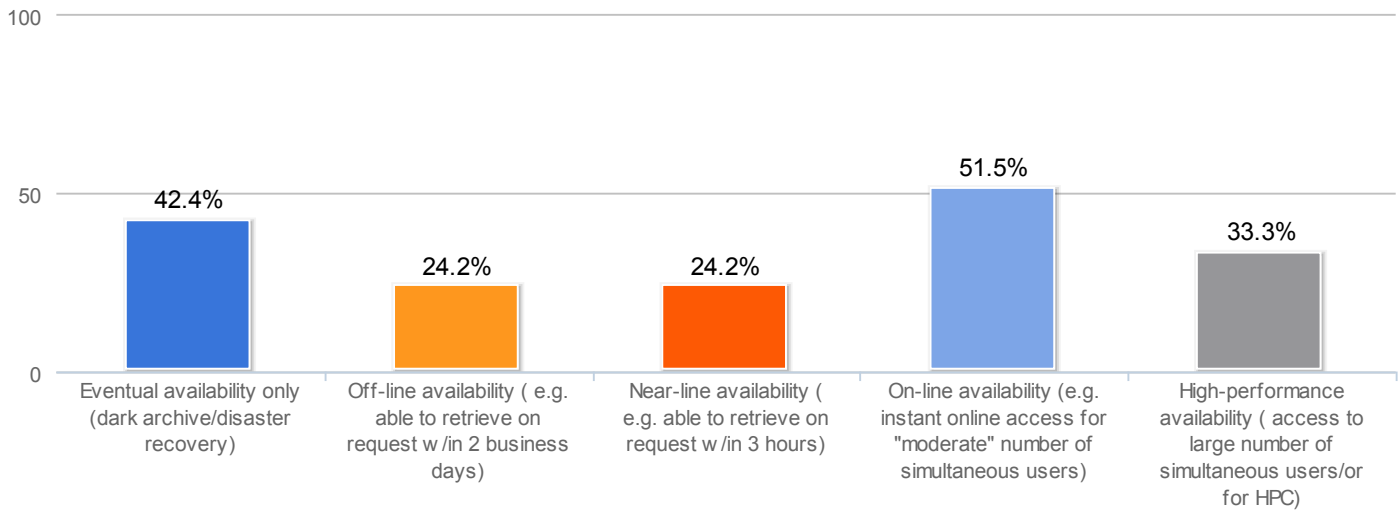

Does your organization have specific documented requirements for your storage system(s) (Check any and all that apply)

Value	Count	Percent %
We have documented general performance requirements	10	33.3%
We have documented performance requirements for ingest	6	20%
We have documented performance requirements for migration to new technology or other one-time intensive operations.	6	20%
We have documented functional requirements	18	60%
We have documented security requirements	15	50%
We plan to develop requirements within one year	6	20%
We have other documented requirements (Specify)	5	16.7%

Statistics	
Total Responses	30

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)
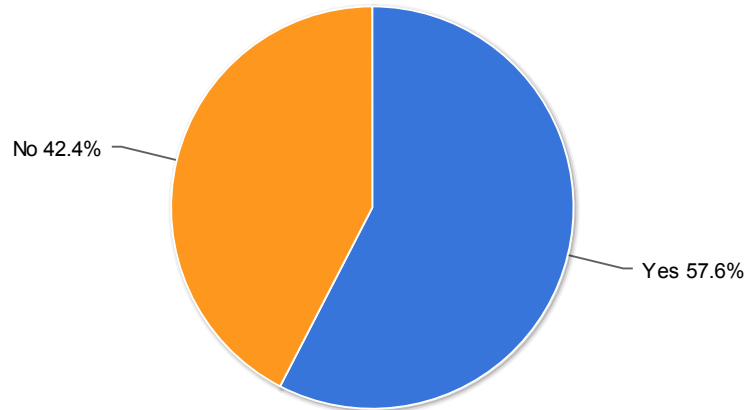

What are your requirements for access to the content you store? (If you have different requirements for different collections please check each option that applies to one of your collections)

Value	Count	Percent %
Eventual availability only (dark archive/disaster recovery)	14	42.4%
Off-line availability (e.g. able to retrieve on request w/in 2 business days)	8	24.2%
Near-line availability (e.g. able to retrieve on request w/in 3 hours)	8	24.2%
On-line availability (e.g. instant online access for "moderate" number of simultaneous users)	17	51.5%
High-performance availability (access to large number of simultaneous users/or for HPC)	11	33.3%

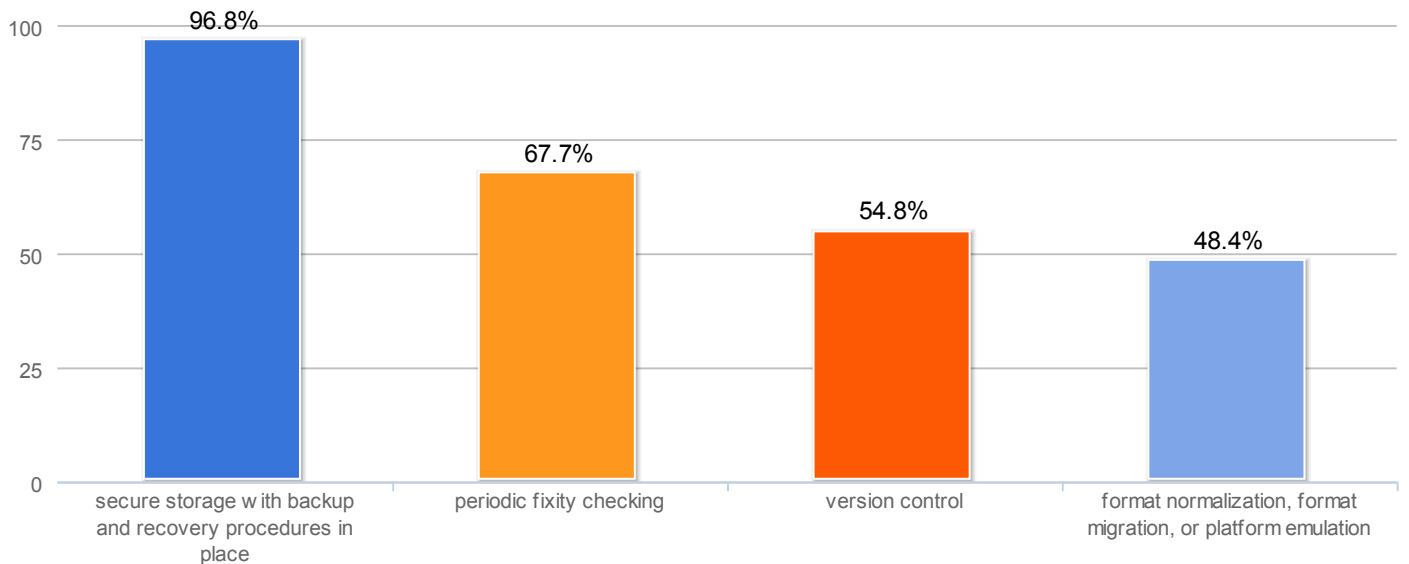| Statistics	
Total Responses	33

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

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

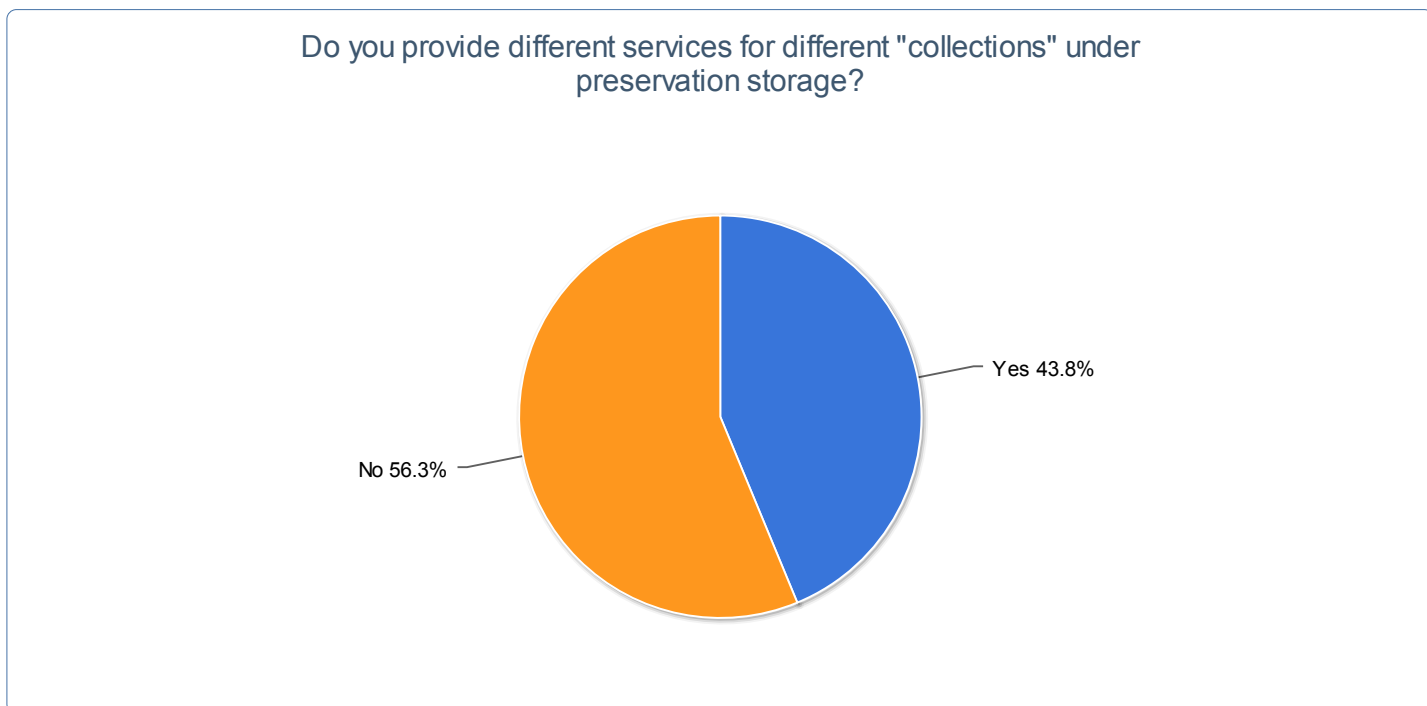| Value	Count	Percent %
Yes	19	57.6%
No	14	42.4%

Statistics	
Total Responses	33

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

Value	Count	Percent %
secure storage with backup and recovery procedures in place	30	96.8%
periodic fixity checking	21	67.7%
version control	17	54.8%
format normalization, format migration, or platform emulation	15	48.4%

Statistics	
Total Responses	31

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

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

Value	Count	Percent %
Yes	14	43.8%
No	18	56.3%

Statistics	
Total Responses	32

If you do provide different services for different collections please describe them below.
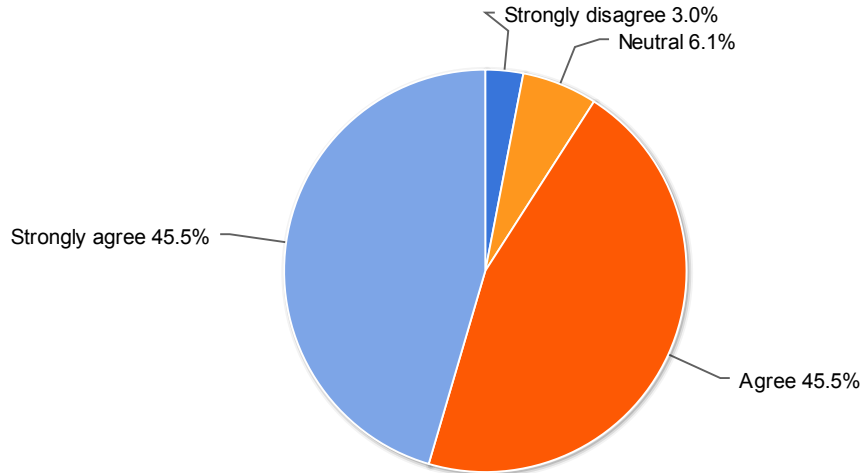
Count	Response
1	No format migration for collection managed by others.
1	Services are determined each collection.
1	We provide differential access services for collections, based on requirements.
1	Under special circumstance, we handle secure materials manually and provide access to only authorized personnel.
1	This is a test operation although internally produced new content is archived in this system. Film

1	This is a test operation although internally produced new content is archived in this system. Film elements, video elements and general high-value content in the Academy Archive and Library currently remain outside of scope during the testing phase.
1	Access services depend on content type. We operate two different access systems, one for Web archives and one general purpose access system for non-Web archive content such as books, maps, photographs, datasets, research papers and newspapers
1	A portion of the content is still managed in an older legacy repository solution with a lower service level (single copy only). We expect to migrate all legacy content to our newer repository over the next year.
1	If a work of digital art doesn't have a specific web page, we work with the artist to determine how to best present their work online.
1	Three broad classes: - Global LOCKSS network - CLOCKSS network - Various Private LOCKSS networks
1	High volume records with selective access restrictions and records in a propritary format are offered are handled in house, unlike the most collections
1	It depends on what is meant by "service". We have different preservation services and they have different business models. In terms of the acts that make up preservation, the steps are similar for all, but do have some differences.

How significant are each of the following general features of preservation systems for meeting your organizations objectives. (1 being insignificant 7 being very significant)

	1 (least)	2	3	4	5	6	7 (most)	Total
More storage	0.0% 0	3.2% 1	9.7% 3	16.1% 5	19.4% 6	25.8% 8	25.8% 8	100% 31
Block level access to storage (Not just file level)	19.4% 6	32.3% 10	6.5% 2	16.1% 5	19.4% 6	6.5% 2	0.0% 0	100% 31
Higher performance processing capacity (to do processing like indexing on content)	0.0% 0	6.5% 2	6.5% 2	9.7% 3	41.9% 13	22.6% 7	12.9% 4	100% 31
More built-in functions (like fixity checking)	0.0% 0	9.4% 3	0.0% 0	12.5% 4	25.0% 8	28.1% 9	25.0% 8	100% 32
More automated inventory, retrieval and management services	0.0% 0	3.1% 1	12.5% 4	15.6% 5	21.9% 7	15.6% 5	31.3% 10	100% 32
More security for the content	3.1% 1	6.3% 2	18.8% 6	21.9% 7	25.0% 8	12.5% 4	12.5% 4	100% 32
File Format migration	3.2% 1	9.7% 3	12.9% 4	12.9% 4	38.7% 12	12.9% 4	9.7% 3	100% 31

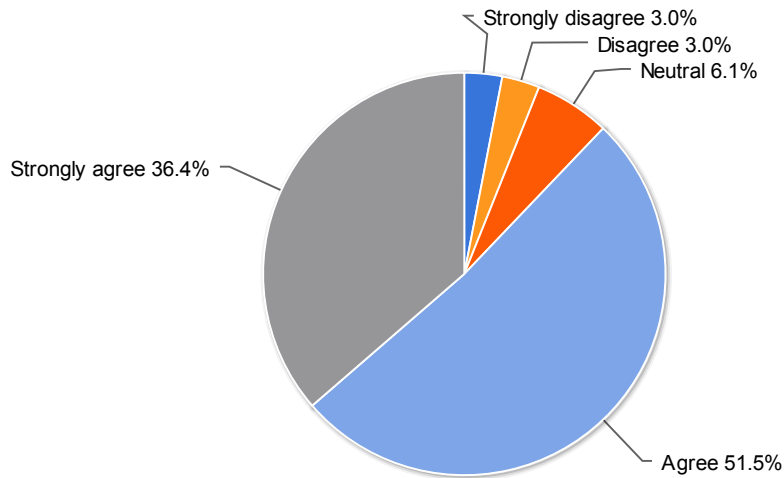## My organization has a plan to meet our preservation storage requirements over the next three years.

Strongly disagree 3.0%
Neutral 6.1%
Strongly agree 45.5%
Agree 45.5%

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

Value	Count	Percent %
Strongly disagree	1	3%
Neutral	2	6.1%
Agree	15	45.5%
Strongly agree	15	45.5%

Statistics	
Total Responses	33
Sum	142.0
Average	4.3
StdDev	0.83
Max	5.0

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

Count	Response
1	5
2	50
2	99
27	999

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

Strongly disagree 3.0%
Disagree 3.0%
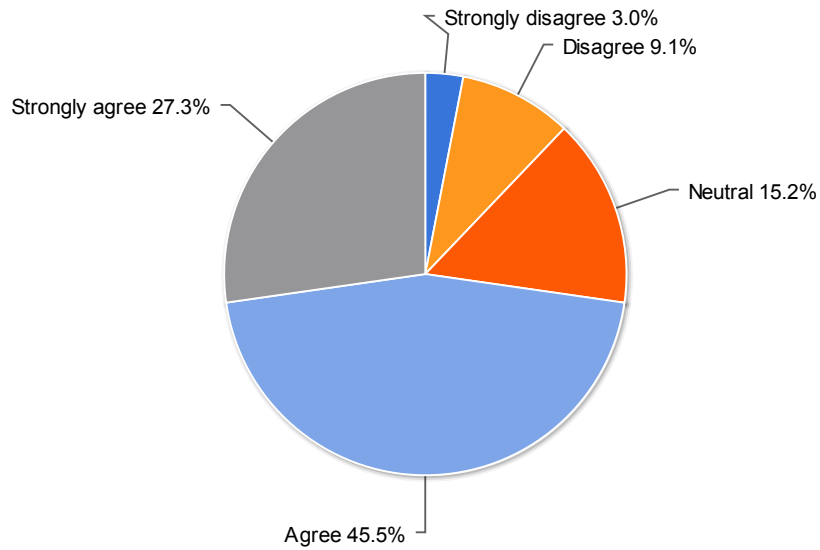Neutral 6.1%
Strongly agree 36.4%
Agree 51.5%

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

Value	Count	Percent %
Strongly disagree	1	3%
Disagree	1	3%
Neutral	2	6.1%
Agree	17	51.5%
Strongly agree	12	36.4%

Statistics	
Total Responses	33
Sum	137.0
Average	4.2
StdDev	0.89
Max	5.0

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

Strongly disagree 3.0%

Disagree 9.1%

Strongly agree 27.3%

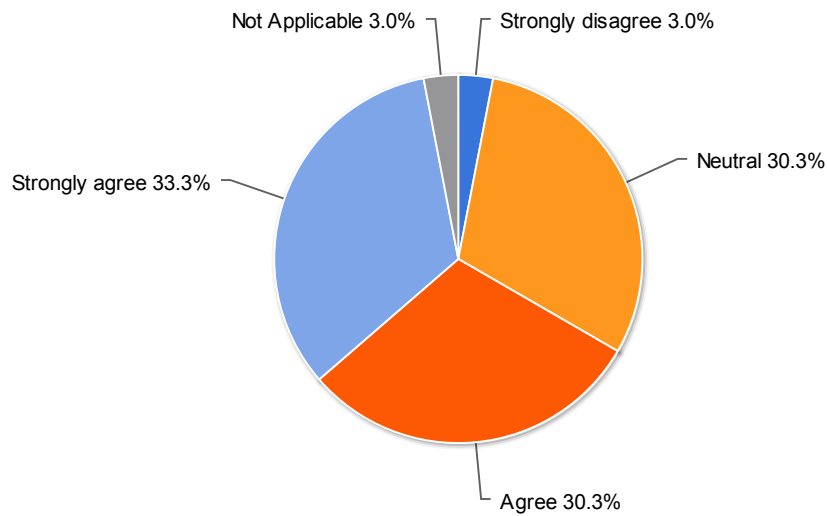Neutral 15.2%

Agree 45.5%

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

Value	Count	Percent %
Strongly disagree	1	3%
Disagree	3	9.1%
Neutral	5	15.2%
Agree	15	45.5%
Strongly agree	9	27.3%

Statistics	
Total Responses	33
Sum	127.0
Average	3.8
StdDev	1.02
Max	5.0

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.

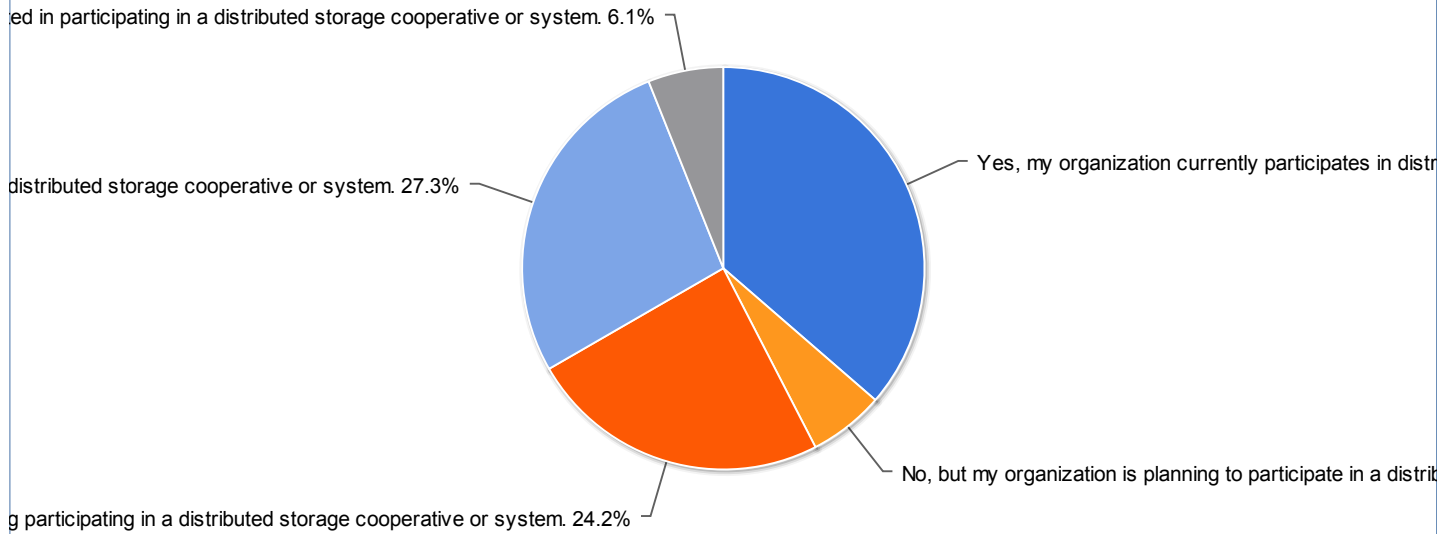## My organization intends to meet requirements for a trustworthy digital repository according to TRAC or the planned ISO standard 16363 within the next three years.

Value	Count	Percent %
Strongly disagree	1	3%
Neutral	10	30.3%
Agree	10	30.3%
Strongly agree	11	33.3%
Not Applicable	1	3%

Statistics	
Total Responses	33
Sum	126.0
Average	3.9
StdDev	0.97
Max	5.0

Is your organization participating in a distributed storage cooperative or system? (ex. LOCKSS alliance, MetaArchive, Data-PASS)

Is your organization participating in a distributed storage cooperative or system? (ex. LOCKSS alliance, MetaArchive, Data-PASS)

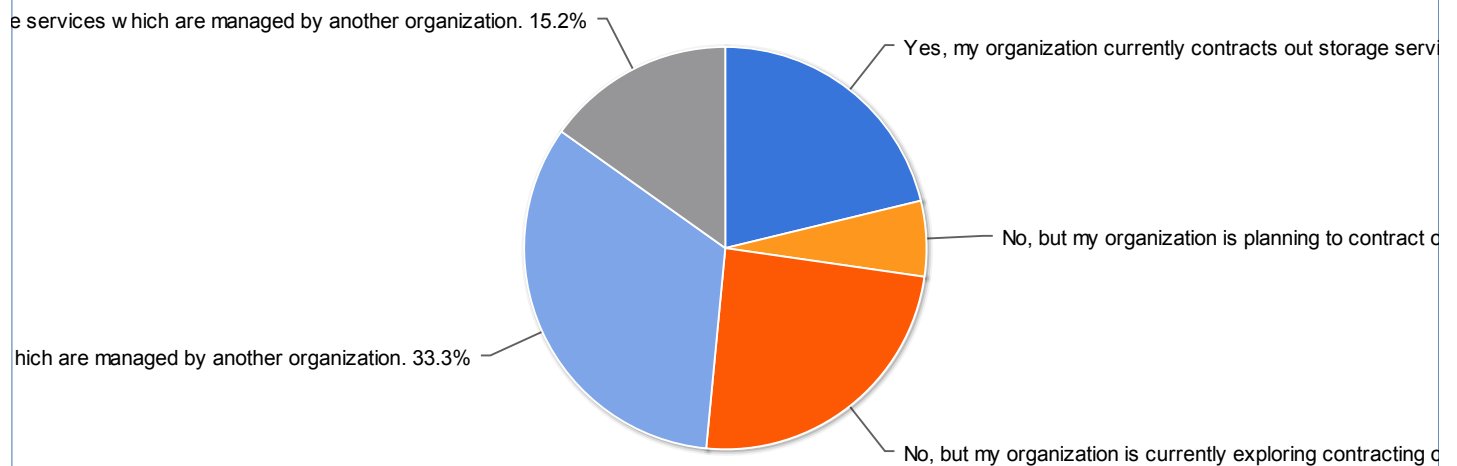| Value	Count	Percent %
Yes, my organization currently participates in distributed storage cooperative or system.	12	36.4%
No, but my organization is planning to participate in a distributed storage cooperative or system.	2	6.1%
No, but my organization is currently exploring participating in a distributed storage cooperative or system.	8	24.2%
No, my organization is not considering participating in a distributed storage cooperative or system.	9	27.3%
No, and my organization is uninterested in participating in a distributed storage cooperative or system.	2	6.1%

Statistics	
Total Responses	33

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

Count	Response
1	Chronopolis
1	Data-PASS
1	Hathi Trust
2	LOCKSS
1	LOCKSS MetaArchive
1	LOCKSS, CLOCKSS, HathiTrust

1	LOCKSS, CLOCKSS, PLNs
1	LOCKSS, Portico
1	MetaArchive
1	Private LOCKSS
1	Working with IRODS on a grant
1	iRods - CineGrid Exchange
1	statewide cooperative
1	We have talked to the Washington State Archives about partnering to provide regional storage services for other state institutions

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

... services which are managed by another organization. 15.2%
Yes, my organization currently contracts out storage serv...
No, but my organization is planning to contract ...
... hich are managed by another organization. 33.3%
No, but my organization is currently exploring contracting ...

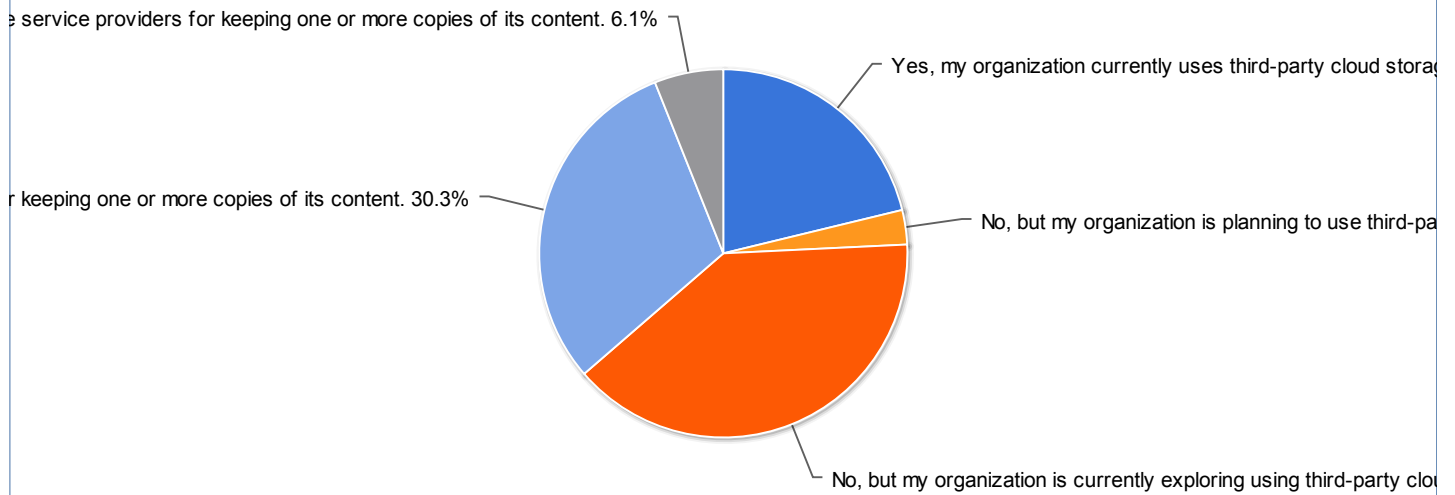**Is your organization contracting out storage services to be managed by another organization or company?**

Value	Count	Percent %
Yes, my organization currently contracts out storage services which are managed by another organization.	7	21.2%
No, but my organization is planning to contract out storage services which are managed by another organization.	2	6.1%
No, but my organization is currently exploring contracting out storage services which are managed by another organization.	8	24.2%
No, my organization is not considering contracting out storage services which are managed by another organization.	11	33.3%
No, and my organization is uninterested in considering contracting out storage services which are managed by another organization.	5	15.2%

Statistics	
Total Responses	33

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.

Count	Response
1	Cloud storage (as a secondary or tertiary copy)
1	OCLC Digital Archive
1	Preservation copies managed at remote sites
1	Rackspace Hosting
1	Some of our infrastructure supporting our replicas is maintained by others.
1	TACC
1	University of Michigan
1	Various S3 "compatible" services
1	We might be interested if there was a viable solution for media.
1	cloud storage vendors
1	offline tape storage, StorageAAS/Cloud

Is your organization using third-party cloud storage service providers (E.g.

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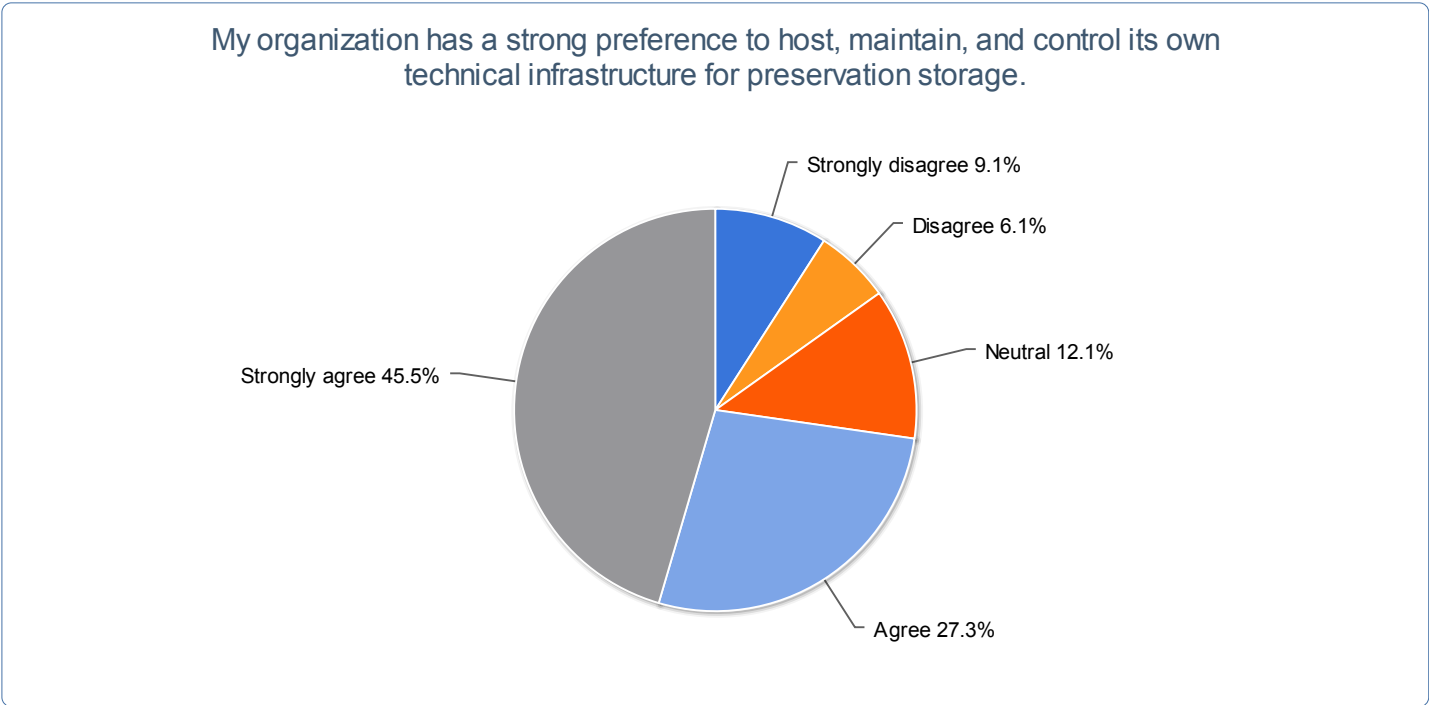

Value	Count	Percent %
Yes, my organization currently uses third-party cloud storage service providers for keeping one or more copies of its content.	7	21.2%
No, but my organization is planning to use third-party cloud storage service providers for keeping one or more copies of its content.	1	3%
No, but my organization is currently exploring using third-party cloud storage service providers for keeping one or more copies of its content.	13	39.4%
No, my organization is not considering using third-party cloud storage service providers for keeping one or more copies of its content.	10	30.3%
No, and my organization is uninterested in using third-party cloud storage service providers for keeping one or more copies of its content.	2	6.1%

Statistics	
Total Responses	33

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

Count	Response
1	AWS GovCloud
1	Amazon S3

1	Amazon, DuraSpace
1	Amazon, Duracloud
1	Amazon, Rackspace
1	Chronopolis
1	DropBox, moving to Rackspace Cloud within 6mos
4	DuraCloud
1	DuraCloud, Amazon, Rackspace
1	DuraCloud, UC San Diego
1	DuraCloud, others TBD
1	Duracloud
1	S3
1	Tested applications on Amazon, Rackspace; watching Duracloud
1	Various S3 "compatible" services
1	box.net
1	too early to say

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

Strongly disagree 9.1%
Disagree 6.1%
Neutral 12.1%
Agree 27.3%
Strongly agree 45.5%

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

Value	Count	Percent %	Statistics

Value	Count	Percent %
Strongly disagree	3	9.1%
Disagree	2	6.1%
Neutral	4	12.1%
Agree	9	27.3%
Strongly agree	15	45.5%

Statistics	
Total Responses	33
Sum	130.0
Average	3.9
StdDev	1.28
Max	5.0

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?

Count	Response
1	Cost, security
1	Desire to maintain control of records that it is mandated to maintain
1	IP
1	NLM is a library of last resort
1	Risk management, policy, data security and cost.
1	Sheer scale of our storage requirements makes it cost prohibitive to outsource fully.
1	TCO
1	To be better able to meet our long-term preservation and curation obligations.
1	We are a technology provider; we need to know how to do it.
1	We can ensure stewardship if we control it.
1	We have the experience for it.
1	We manage Chronopolis, which provides preservation management.
1	bandwidth management
1	cost, security, government requirements
1	Lack of trust. Lack of standards. Not sure of implications for trustworthiness. Not sure it can be done better and cheaper than we do.
1	Prefer to control our own infrastructure at this time. This could change at some point in the future.
1	This has been the traditional stance, although there is a trend to move larger projects to cloud storage providers as long as they are self sustainable.
1	Government has a responsibility to maintain the records of the people and should not abdicate this responsibility. It is my belief that is far more cost effective over the long run for government to manage its own record resources.
1	State-consolidated IT offers services but storage prices are too high, so management wants to go alone, whether that means keeping our own hardware in our own building, or moving to the cloud. Whether or not IT will go along with that plan is still an issue. In the past, they have had veto power.
1	For quick access, and because no real viable solutions for large media files other than an in house system.
1	Ideally, we would like to be in control of everything that happens with the content and exactly how it is

managed, but because we are not capable of doing this we replicate our local storage at a vendor as well.

Please provide any additional comments or clarifications here

Count	Response
1	Estimate of storage in T is a guess....will try to confirm.
1	Issues around #22 are mostly about sustainability.
1	This survey has been completed from an internal DuraSpace perspective: policies over our internal content/collection; not from the perspective of the archiving/preservation services we offer via DuraCloud, etc. There is significant overlap in what we do internally and what we offer; but total storage needs (1-TB) and storage in a distributed cooperative system, are examples that reflect this difference.
1	We are just beginning to have applications in place to manage ingest and similar features, so while format transformation/migration is not yet available, it soon should be. The same goes with automatic checksums. We are very busy with system design work.
1	We are still exploring possible architectures and systems. Strong Preference is not indicated either way at this point.
1	Note: our backup (tape) copies are stored offsite by a third party. Question 19: The Library does have a LOCKSS box, but it is not managed by my department.

My NDSA member organization is ...

Count	Response
1	AMPAS
1	CIESIN, Columbia University
1	California Digital Library
1	DuraSpace
1	FCLA
1	Harvard
1	Harvard Library
1	ICPSR
1	Internet Archive
1	Johns Hopkins Sheridan Libraries
1	KY Dept. for Libraries and Archives
1	LOCKSS
1	Maryland State Archives
1	Michigan State University Library

1	NC Department of Cultural Resources
2	National Library of Medicine
1	New York University
1	Odum institute
1	Penn State
1	Portico
1	Rhizome
1	Roy Rosenzweig Center for History and New Media
1	Stanford
1	UC San Diego
1	United States Holocaust Memorial Museum
1	University of Minnesota Libraries
1	University of North Texas
1	University of Southern California
1	Utah State Archives
1	WGBH Educational Foundation
1	i/oTrak E-Z Photo Scan