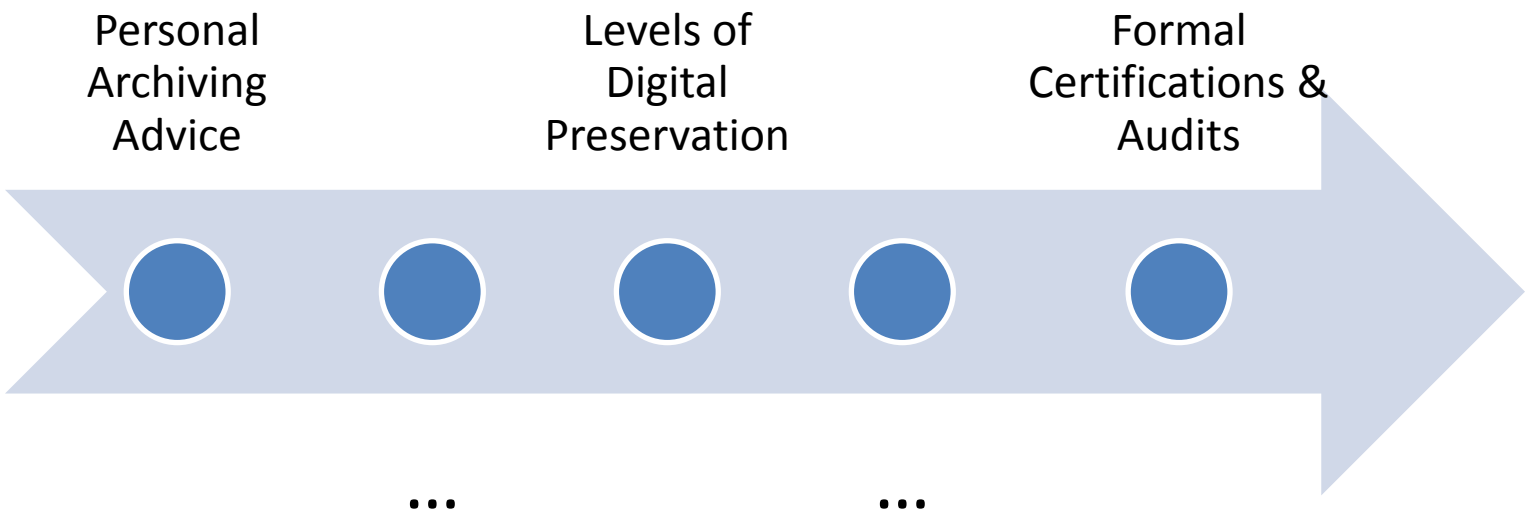
NDSA Levels of Digital Preservation Update

Common Need

- Simple, practical, documented levels of preservation services reflecting best practices, **broadly useful**
  - For those just starting out & those with mature programs
  - Independent of formats, storage systems
  - Useful to educators & implementers

Niche


Personal Archiving Advice
Levels of Digital Preservation
Formal Certifications & Audits
...
...

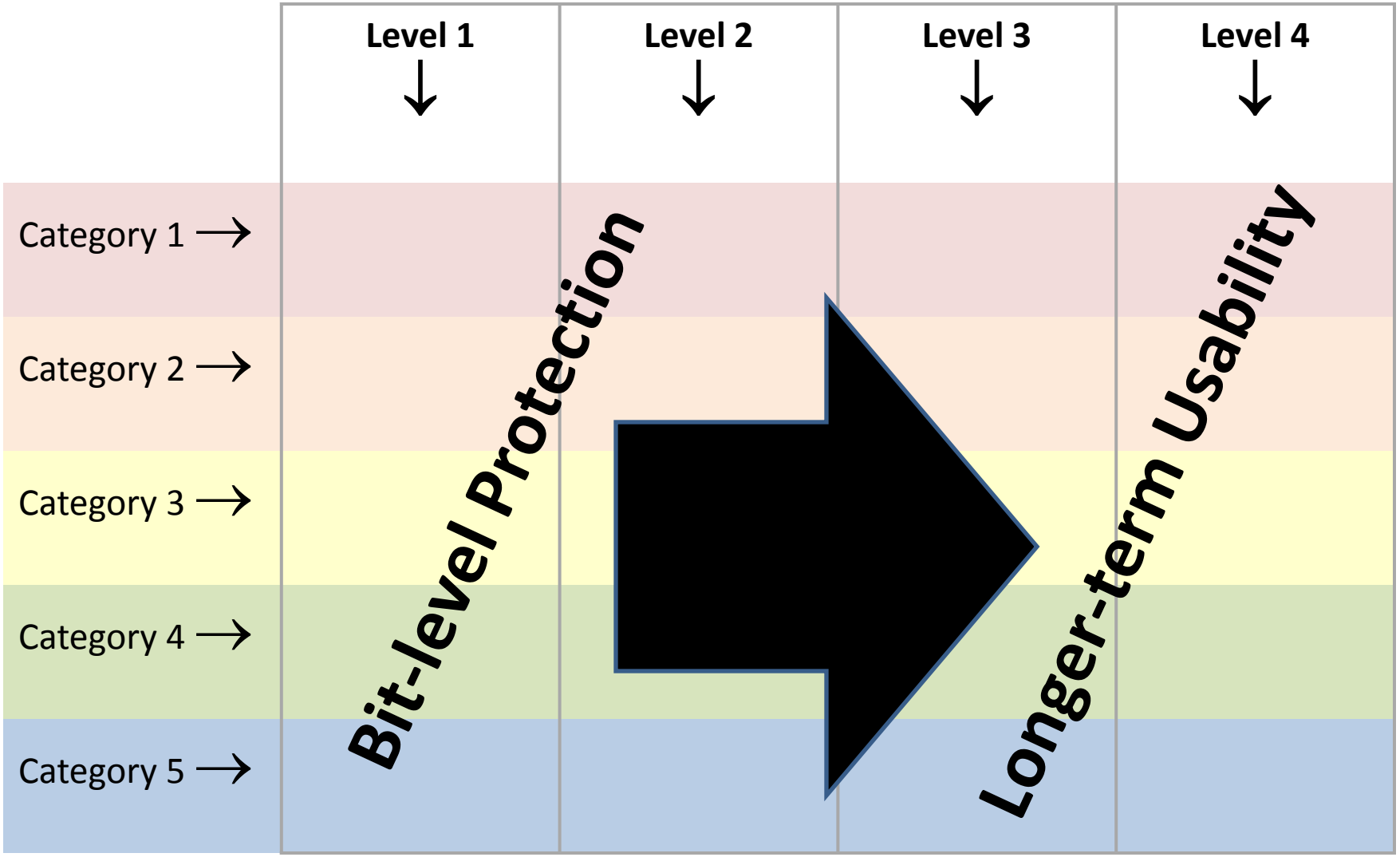
Levels of Digital Preservation, v1

	Level 1 ↓	Level 2 ↓	Level 3 ↓	Level 4 ↓
Category 1 →				
Category 2 →				
Category 3 →				
Category 4 →				
Category 5 →				

Levels of Digital Preservation, v1

	Level 1 ↓	**Level 2** ↓	**Level 3** ↓	**Level 4** ↓
Category 1 →	Level 1 Actions for Category 1	Level 2 Actions for Category 1	…	…
Category 2 →	Level 1 Actions for Category 2	Level 2 Actions for Category 2	…	…
Category 3 →	…	…	…	…
Category 4 →	…	…	…	…
Category 5 →	…	…	…	…

Levels of Digital Preservation, v1



Levels of Digital Preservation, v1

	Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
Storage and Geographic Location	- Two complete copies that are not collocated - For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	- At least three complete copies - At least one copy in a different geographic location - Document your storage system(s) and storage media and what you need to use them	- At least one copy in a geographic location with a different disaster threat - Obsolescence monitoring process for your storage system(s) and media	- At least three copies in geographic locations with different disaster threats - Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	- Check file fixity on ingest if it has been provided with the content - Create fixity info if it wasn't provided with the content	- Check fixity on all ingests - Use write-blockers when working with original media - Virus-check high risk content	- Check fixity of content at fixed intervals - Maintain logs of fixity info; supply audit on demand - Ability to detect corrupt data - Virus-check all content	- Check fixity of all content in response to specific events or activities - Ability to replace/repair corrupted data - Ensure no one person has write access to all copies
Information Security	- Identify who has read, write, move and delete authorization to individual files - Restrict who has those authorizations to individual files	- Document access restrictions for content	- Maintain logs of who performed what actions on files, including deletions and preservation actions	- Perform audit of logs
Metadata	- Inventory of content and its storage location - Ensure backup and non-collocation of inventory	- Store administrative metadata - Store transformative metadata and log events	- Store standard technical and descriptive metadata	- Store standard preservation metadata
File Formats	- When you can give input into the creation of digital files encourage use of a limited set of known open formats and codecs	- Inventory of file formats in use	- Monitor file format obsolescence issues	- Perform format migrations, emulation and similar activities as needed

Storage and Geographic Location

Level 1 Protect your data	Level 2 Know your data	Level 3 Monitor your data	Level 4 Repair your data
Two complete copies that are not collocated For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system	At least three complete copies At least one copy in a different geographic location Document your storage systems(s) and storage media and what you need to use them	At least one copy in a geographic location with a different disaster threat Obsolescence monitoring for your storage system(s) and media	At least three copies in geographic locations with different disaster threats Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems

File Fixity and Data Integrity

Level 1 Protect your data	Level 2 Know your data	Level 3 Monitor your data	Level 4 Repair your data
Check file fixity on ingest if it has been provided with the content Create fixity info if it wasn't provided with the content	Check fixity on all ingests Use write-blockers when working with original media Virus-check high risk content	Check fixity of content at fixed intervals Maintain logs of fixity info; supply audit on demand Ability to detect corrupt data Virus-check all content	Check fixity of all content in response to specific events or activities Ability to replace/repair corrupted data Ensure no one person has write access to all copies

Information Security

Level 1 Protect your data	Level 2 Know your data	Level 3 Monitor your data	Level 4 Repair your data
Identify who has read, write, move and delete authorization to individual files Restrict who has those authorizations to individual files	Document access restrictions for content	Maintain logs of who performed what actions on files, including deletions and preservation actions	Perform audit of logs

Metadata

Level 1 Protect your data	Level 2 Know your data	Level 3 Monitor your data	Level 4 Repair your data
Inventory of content and its storage location Ensure backup and non-collocation of inventory	Store administrative metadata Store transformative metadata and log events	Store standards technical and descriptive metadata	Store standard preservation metadata

File Formats

Level 1 Protect your data	Level 2 Know your data	Level 3 Monitor your data	Level 4 Repair your data
When you can give input into the creation of digital files, encourage use of a limited set of known open formats and codecs	Inventory of file formats in use	Monitor file format obsolescence issues	Perform format migrations, emulation and similar activities as needed

Preliminary Results

2013 NDSA Storage Survey

Usage Contexts

- **Inform Local Guidelines Development:** Educate and develop guidelines for content creators and contributors **USGS**
- **Self Assessments** – how do we compare with best practices? What should we improve next? Where do we excel? How will we improve after project X? How have we improved over time? **Harvard & ARTstor**
- **Developing requirements** for third-party preservation service providers

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

Answer	Response	%
Yes	54	64%
No	30	36%

Project of the NDSA Infrastructure Working Group

The Infrastructure Working Group works to identify and share 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.

My organization's preservation storage system uses the following media.

Answer	Response	%
Spinning disk - Locally or network attached storage (NAS)	57	68%
Spinning disk - Storage area network (SAN)	43	51%
Magnetic tape	40	48%
Other (Please specify)	14	17%

Is your organization keeping copies of digital assets in geographically distinct places

Answer	Response	%
Yes, we manage our own copies in one or more geographically distinct offsite locations	36	43%
Yes, we keep additional copies of our materials in a distributed collaborative partnership, (E.g. MetaArchive)	8	10%
Yes, we keep one or more additional copies of our materials managed by another institution or commercial provider	21	25%
In some cases, decided on collection basis	15	18%
No, we would like to but we do not have the resources	17	20%
No, we do not and this is not something we are pursuing	6	7%

Does your organization follow practices with respect to fixity checking?

Answer	Response	%
Yes, we do fixity checks before and after transactions like ingest	37	44%
Yes, we do fixity checks on all content we are preserving at fixed intervals (E.g. every 9 months)	21	25%
Yes, we randomly sample content and check for fixity	21	25%
Yes, we store fixity information in an independent system	15	18%
Yes, we use a tamper-resistant fixity check mechanism (E.g. LOCKSS, ACE)	14	17%
No, we do not do fixity checks on our content	29	35%

Does your organization have documented requirements for your preservation storage

Answer	Response	%
Yes, we have documented general performance requirements	23	28%
Yes, we have documented performance requirements for ingest	12	15%
Yes, we have documented performance requirements for migration to new technology or other one-time intensive operations.	6	7%
Yes, we have documented functional requirements	26	32%
Yes, we plan to develop requirements within one year	13	16%
Yes, we have other documented requirements (Please specify)	13	16%
No, we do not have documented requirements, but plan to develop requirements within one year	19	23%
No, we do not have documented requirements, and do not plan to develop them	9	11%

What are your organization's requirements for availability of the content you store?

Answer	Response	%
Eventual availability only (dark archive/disaster recovery)	24	29%
Off-line availability (e.g. able to retrieve on request w/in 2 business days)	32	39%
Near-line availability (e.g. able to retrieve on request w/in 3 hours)	25	30%
On-line availability (e.g. instant online access for "moderate" number of simultaneous users)	47	57%
High-performance availability (access to large number of simultaneous users/or for HPC)	13	16%
Not applicable	2	2%

Which services does your organization currently provide for files in your preservation storage?

Answer	Response	%
Secure storage with backup and recovery procedures in place	68	87 %
Periodic fixity checking	40	51 %
Version control	30	38 %
Format normalization, format migration, or platform emulation	23	29 %
Other services (Please describe)	12	15 %
None	5	6%

Does your organization provide different services for different "collections" under preservation

Answer	Response	%
Yes	35	43%
No	47	57%

How significant are each of the following general features of preservation systems for meeting your goals?

Answer (1 most important 7 least)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
More storage	38	7	11	6	2	4	7
More built-in functions (like fixity checking)	9	22	15	10	10	5	4
More automated inventory, retrieval and management services	9	19	15	14	11	5	2
More security for the content	8	7	14	15	16	12	3
File format migration	6	9	8	11	19	13	9
Higher performance processing capacity (to do processing like indexing on content)	4	8	12	16	13	17	5
Block level access to storage (Not just file level)	1	3	0	3	4	19	45

My organization has a plan to meet our preservation storage requirements over the next 3 years.

Answer	Response	%
Strongly Disagree	2	3%
Disagree	8	11%
Neutral	19	25%
Agree	29	38%
Strongly Agree	18	24%

My organization plans to make significant changes in technologies in its preservation storage ar...

Answer	Response	%
Strongly Disagree	1	1%
Disagree	11	14%
Neutral	21	28%
Agree	27	36%
Strongly Agree	16	21%

My organization intends to meet the requirements for a trustworthy digital repository

Answer	Response	%
Strongly Disagree	5	7%
Disagree	10	13%
Neutral	25	33%
Agree	24	32%
Strongly Agree	12	16%

For those intending to meet a standard, which digital repository standard(s) your organization is targeting

Answer	Response	%
ISO 16363	18	56%
TRAC	24	75%
Data Seal of Approval	7	22%
Other (Please specify)	4	13%

23. 18. Is your organization participating in a distributed storage cooperative or system? (ex. LOCKSS A...

Answer	Response	%
Yes, my organization currently participates in a distributed storage cooperative or system.	28	39%
No, but my organization is planning to participate in a distributed storage cooperative or system.	1	1%
No, but my organization is currently exploring participating in a distributed storage cooperative or system.	20	28%
No, my organization is not considering participating in a distributed storage cooperative or system.	18	25%
No, and my organization is uninterested in participating in a distributed storage cooperative or system.	4	6%

Excluding distributed storage cooperatives, is your organization using third-party cloud storage?

Answer	Response	%
Yes, my organization currently uses third-party cloud storage service providers for keeping one or more copies of its content.	17	23%
No, but my organization is planning to use third-party cloud storage service providers for keeping one or more copies of its content.	12	16%
No, but my organization is currently exploring using third-party cloud storage service providers for keeping one or more copies of its content.	25	34%
No, my organization is not considering using third-party cloud storage service providers for keeping one or more copies of its content.	15	21%
No, and my organization is uninterested in using third-party cloud storage service providers for keeping one or more copies of its content.	4	5%

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

Answer	Response	%
Strongly Disagree	5	7%
Disagree	8	11%
Neutral	24	32%
Agree	17	23%
Strongly Agree	21	28%