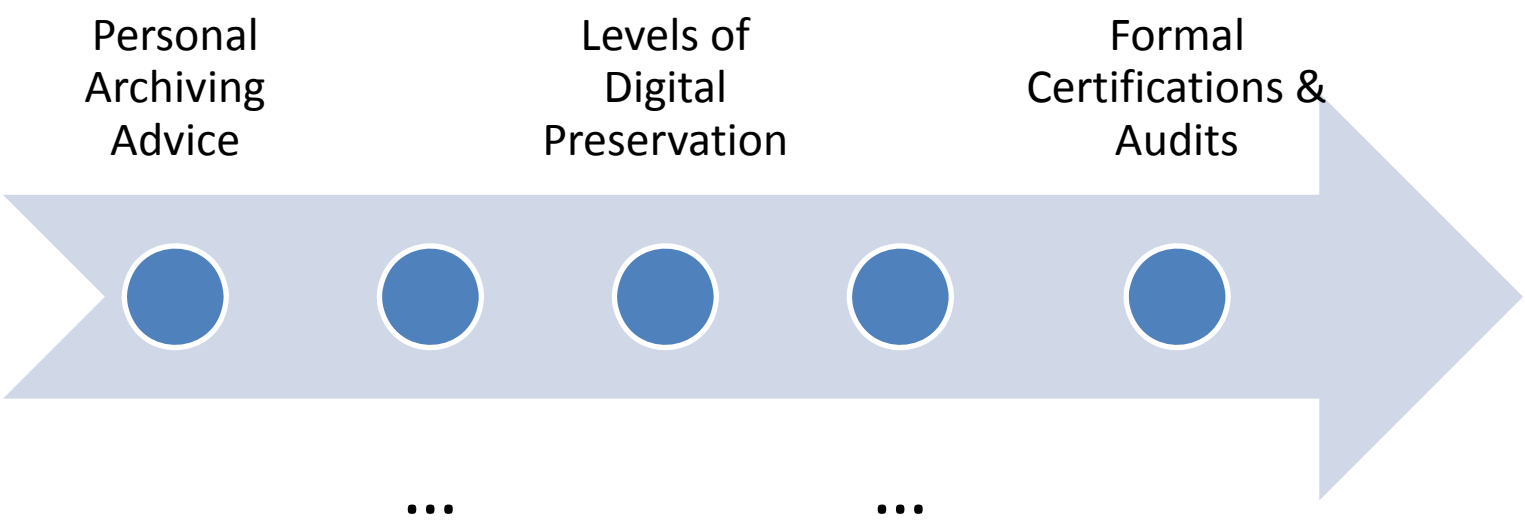
# NDSA Levels of Digital Preservation Update

# 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
...
...

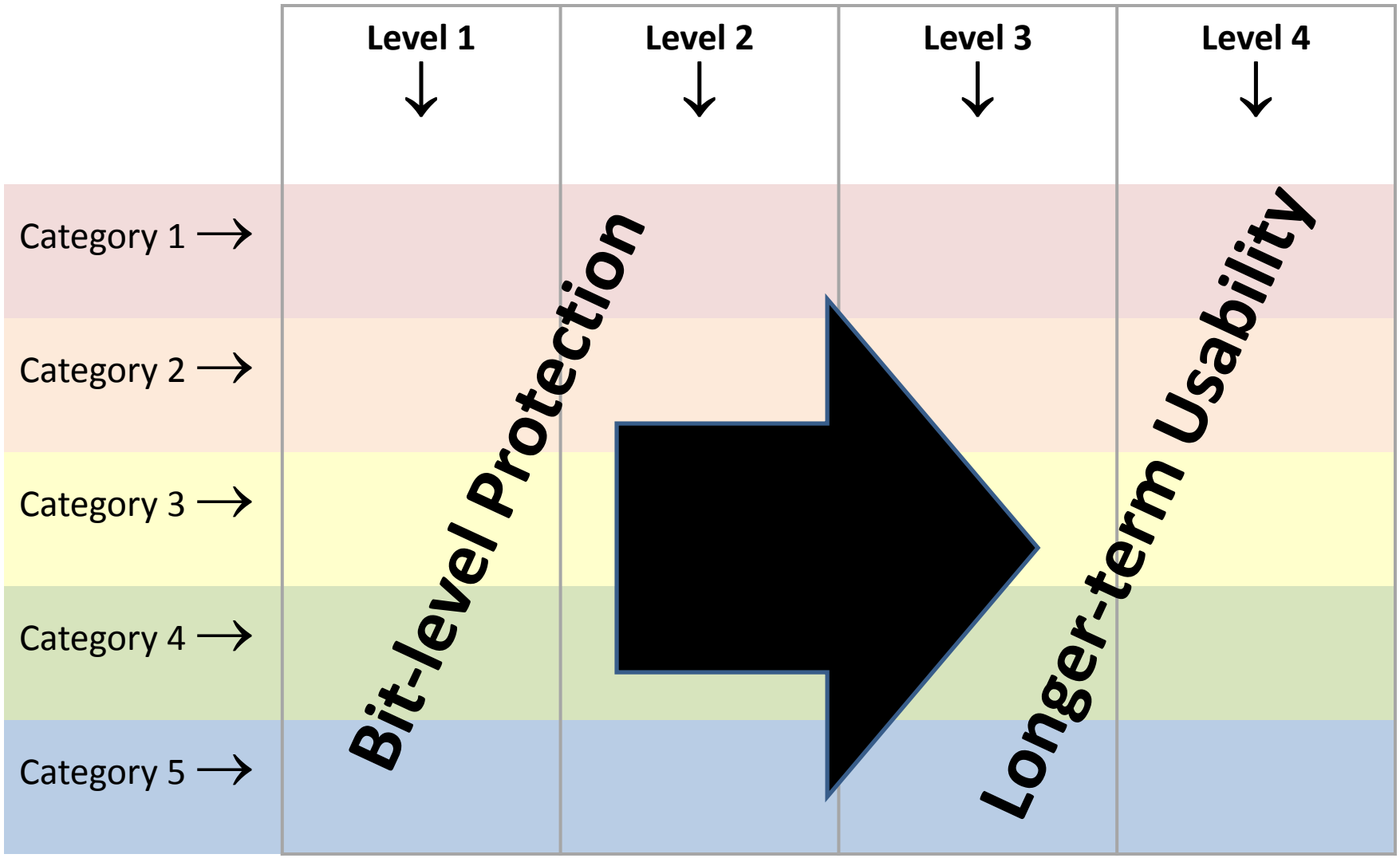
# Levels of Digital Preservation, v1

| | 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<br>- 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<br>- At least one copy in a different geographic location<br>- 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<br>- Obsolescence monitoring process for your storage system(s) and media | - At least three copies in geographic locations with different disaster threats<br>- 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<br>- Create fixity info if it wasn't provided with the content | - Check fixity on all ingests<br>- Use write-blockers when working with original media<br>- Virus-check high risk content | - Check fixity of content at fixed intervals<br>- Maintain logs of fixity info; supply audit on demand<br>- Ability to detect corrupt data<br>- Virus-check all content | - Check fixity of all content in response to specific events or activities<br>- Ability to replace/repair corrupted data<br>- Ensure no one person has write access to all copies |
| **Information Security** | - Identify who has read, write, move and delete authorization to individual files<br>- 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<br>- Ensure backup and non-collocation of inventory | - Store administrative metadata<br>- 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<br>Protect your data | Level 2<br>Know your data | Level 3<br>Monitor your data | Level 4<br>Repair your data |
|---|---|---|---|
| Two complete copies that are not collocated<br><br>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<br><br>At least one copy in a different geographic location<br><br>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<br><br>Obsolescence monitoring for your storage system(s) and media | At least three copies in geographic locations with different disaster threats<br><br>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**<br>**Protect your data** | **Level 2**<br>**Know your data** | **Level 3**<br>**Monitor your data** | **Level 4**<br>**Repair your data** |
|---|---|---|---|
| Check file fixity on ingest if it has been provided with the content<br><br>Create fixity info if it wasn't provided with the content | Check fixity on all ingests<br><br>Use write-blockers when working with original media<br><br>Virus-check high risk content | Check fixity of content at fixed intervals<br><br>Maintain logs of fixity info; supply audit on demand<br><br>Ability to detect corrupt data<br><br>Virus-check all content | Check fixity of all content in response to specific events or activities<br><br>Ability to replace/repair corrupted data<br><br>Ensure no one person has write access to all copies |

# Information Security

| Level 1<br>Protect your data | Level 2<br>Know your data | Level 3<br>Monitor your data | Level 4<br>Repair your data |
|---|---|---|---|
| Identify who has read, write, move and delete authorization to individual files<br><br>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<br>Protect your data | Level 2<br>Know your data | Level 3<br>Monitor your data | Level 4<br>Repair your data |
|---|---|---|---|
| Inventory of content and its storage location<br><br>Ensure backup and non-collocation of inventory | Store administrative metadata<br><br>Store transformative metadata and log events | Store standards technical and descriptive metadata | Store standard preservation metadata |

# File Formats

| Level 1<br>Protect your data | Level 2<br>Know your data | Level 3<br>Monitor your data | Level 4<br>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

# 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.

# Respondents

Participants from 81 member organizations from the NDSA responded to the Survey. These include, libraries, archives, museums, service providers and a range of commercial content holders.

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

| Answer | Response | % |
|---|---|---|
| Yes | 76 | 94% |
| No | 5 | 6% |

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

| Answer | Response | % |
|---|---|---|
| Spinning disk - Locally or network attached storage (NAS) | 54 | 67% |
| Spinning disk - Storage area network (SAN) | 42 | 51% |
| Magnetic tape | 39 | 48% |
| Cloud Storage | 6 | 7% |
| Optical Media | 4 | 5% |
| Other | 3 | 4% |

# Does your organization follow practices with respect to fixity checking?

| Answer | Response | % |
|---|---|---|
| Yes, we do fixity checks before and after transactions like ingest | 36 | 44% |
| Yes, we do fixity checks on all content we are preserving at fixed intervals (E.g. every 9 months) | 23 | 25% |
| Yes, we randomly sample content and check for fixity | 17 | 25% |
| Yes, we store fixity information in an independent system | 17 | 18% |
| Yes, we use a tamper-resistant fixity check mechanism (E.g. LOCKSS, ACE) | 13 | 17% |

# Does your organization have documented requirements for your preservation storage

| Answer | Response | % |
|---|---|---|
| Yes, we have documented functional requirements | 26 | 32% |
| Yes, we have documented general performance requirements | 23 | 28% |
| Yes, we have other documented requirements | 14 | 16% |
| Yes, we have documented performance requirements for migration to new technology or other one-time intensive operations. | 8 | 7% |
| 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) | 25 | 31% |
| On-line availability (e.g. instant online access for "moderate" number of simultaneous users) | 24 | 30% |
| Off-line availability ( e.g. able to retrieve on request w/in 2 business days) | 17 | 21% |
| Near-line availability ( e.g. able to retrieve on request w/in 3 hours) | 9 | 11% |
| High-performance availability ( access to large number of simultaneous users/or for HPC) | 5 | 6% |
| Not applicable | 1 | 1% |

# 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 | 8 | 11 | 8 | 2 | 5 | 7 |
| More built-in functions (like fixity checking) | 11 | 22 | 14 | 11 | 10 | 7 | 4 |
| More automated inventory, retrieval and management services | 9 | 20 | 17 | 14 | 12 | 6 | 1 |
| More security for the content | 8 | 7 | 14 | 18 | 16 | 14 | 2 |
| File format migration | 7 | 10 | 9 | 11 | 20 | 10 | 12 |
| Higher performance processing capacity (to do processing like indexing on content) | 5 | 10 | 14 | 14 | 14 | 17 | 5 |
| Block level access to storage (Not just file level) | 1 | 2 | 0 | 3 | 4 | 20 | 48 |

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

| Answer | Response | % |
|---|---|---|
| Strongly Disagree | 3 | 3% |
| Disagree | 8 | 11% |
| Neutral | 19 | 25% |
| Agree | 31 | 38% |
| Strongly Agree | 20 | 24% |

My organization plans to make significant changes in technologies in its preservation storage architecture in the next 3 years.

| Answer | Response | % |
|---|---|---|
| Strongly Disagree | 2 | 2% |
| Disagree | 13 | 16% |
| Neutral | 20 | 25% |
| Agree | 26 | 32% |
| Strongly Agree | 27 | 21% |

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

| Answer | Response | % |
|---|---|---|
| Disagree | 20 | 25% |
| Neutral | 24 | 32% |
| Agree | 39 | 48% |

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

| Answer | Response |
|---|---|
| ISO 16363 | 21 |
| TRAC | 26 |
| Data Seal of Approval | 8 |
| Other (Please specify) | 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 | 27 | 32% |
| Agree | 19 | 23% |
| Strongly Agree | 22 | 28% |