Third time’s the charm!

Introducing Version 3 of the Digital Preservation Storage Criteria

Sibyl Schaefer presenting for the Digital Preservation Storage Criteria Team

“Preservation Storage”

- Storage that supports digital preservation (“the series of managed activities necessary to ensure continued access to digital materials for as long as necessary” - Digital Preservation Coalition)
- Scoped to cover the combination of:
  - All functions of the OAIS functional entity “archival storage”
  - Some functions of other OAIS functional entities needed to store, maintain in storage, and retrieve from storage
  - Expectations of modern storage solutions
  - Expectations of modern digital preservation practices

Why Needed?

All digital preservation activities rely on storage, yet there are no community guidelines available to aid storage selection.

Potential Uses

- To inform more detailed requirements for preservation storage
- To identify gap areas in existing preservation storage
- To evaluate or compare preservation storage solutions
- To seed discussions with IT about preservation storage
- To seed discussions within the digital preservation field on preservation storage
- As a component of instructional material on digital preservation

Intended Audience

- Providers of all or parts of a Digital Preservation Storage solution
- Consumers of all or parts of a Digital Preservation Storage solution, ranging from institutions just starting out with preservation to institutions with established preservation programs

Guiding Principles

- The Criteria should describe characteristics of preservation storage relevant to a wide range of different kinds of institutions with responsibility for preserving digital material, and to organizations providing preservation storage services to other institutions.
- The Criteria omits text that assumes specific architecture, technology, media, content, policy or vendor choices.
- *Not all of the Criteria will be applicable to all institutions.*

Guiding Principles (cont.)

- The Criteria are not intended to be detailed enough to use as a preservation storage requirements document.
- The Criteria are meant to be used as a foundation for informing preservation storage and to be combined with local policies, applicable regulations, needs and preferences.
- The Criteria do not cover any additional infrastructure needed in combination with preservation storage, e.g. staging areas, testing infrastructure, delivery and management servers.

How are the criteria presented?

- Separated into eight different categories for ease of use.
- Structured list of 61 criteria, each with a short description and reference/citation as available or appropriate.
- Accompanied by a Criteria Usage Guide with supporting materials on risk, cost, and independence factors.

How are the criteria presented?

26	Virus/malware detection	Information security	Includes software that regularly runs virus checks and malware detection.	
27	Virus/malware remediation	Information security	Provides remediation actions for content with viruses and/or malware, e.g. quarantine, notification, etc.	
28	Diverse storage media types	Resilience	Uses different storage media types / configurations / providers together so that desired levels of independence can be achieved	(DP Storage WG, 2018, Independence section)
29	Durable media	Resilience	Provides documented and acceptable longevity, failure rates, and technical characteristics of the storage media components	
30	Error control	Resilience	Performs error detection and correction 24/7/365 (e.g. using RAID, Erasure coding, ZFS, triple copies/rebuild)	
31	High availability	Resilience	Has a high percentage of uptime, i.e. operational for a long length of time, due to techniques such as eliminating single points of failure by having redundant equipment, load-balanced systems and effective monitoring to detect software or hardware failures	(SNIA, 2017)

Evolution of the Criteria

	2016 - Version 1	2017 - Version 2	2018 - Version 3
# Criteria/Categories:	48 / NA	58 / 8	61 / 8
Categories:	None	Content Integrity (3)	Content Integrity (2)
		Cost Considerations (3)	Cost Considerations (3)
		Flexibility & Resiliency (12)	**Flexibility** (7)
		Information Security (11)	Information Security (15)
		Scalability & Performance (11)	Scalability & Performance (10)
		Support (3)	Support (4)
		Transparency (11)	Transparency (14)
		Storage Location (4)	**Resilience** (6)

Evolution of the Criteria

2017 Version 2 - 58 total, 10 new from Version 1:

1. Virus (or malware) checking
2. Logging
3. Authenticity
4. Transparency (of self-healing repair)
5. Environmental Impact
6. Weight
7. Ability to export data out at a reasonable/negotiated rate, exit strategies
8. I/O performance
9. Compute power
10. Quality of storage medium

2018 Version 3 - 61 total:

Normalised criteria wording for consistency

Mapping to relevant standards

What Changed - Examples

Criteria changed categories and renamed:

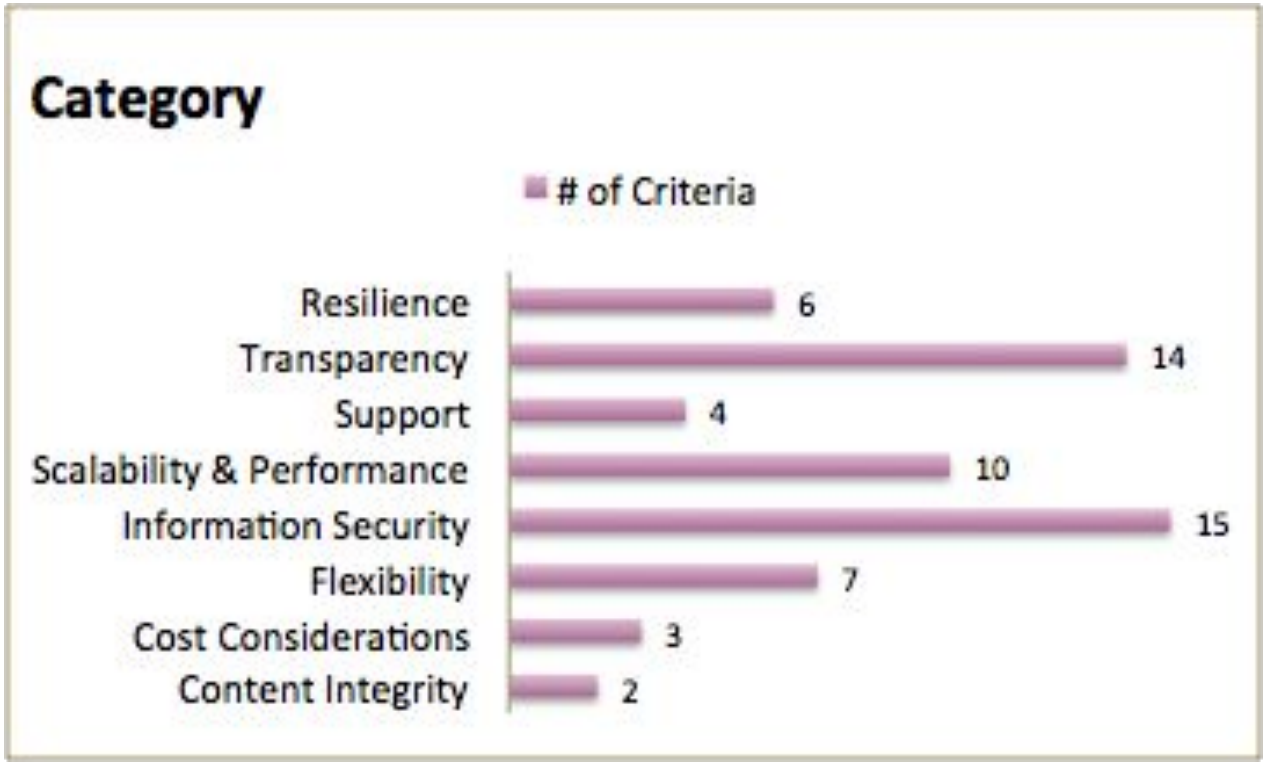
- [Content Integrity, #3] “Provides preservation actions”

→ [Support, #45] “Independent preservation services”

Criteria consistency

- Short description start with actions, .e.g “provides”, “supports”, “allows”
- Removal of institutional-specific text

Version 3 Visual Breakdown



Criteria Usage Guide

- Supplements the Criteria by clarifying and contextualizing the criteria, providing additional considerations, and showing examples of how they may be used.
- Sections:
  - Establishing Bit Integrity
  - Independence Between Copies
  - Cost Analysis
  - Risk Management

Elements in Establishing Bit Integrity

- Number of copies
  - Do we have enough to survive loss(es)?
  - Do we have enough to verify the integrity of other copies?
- Independence of copies (next slide)
- Integrity check (of and between copies)
  - Is our local copy correct according to our recorded fixity values?
  - Is our local copy the same as other distributed copies?

Independence of Copies

- Organizational
  E.g. Decision level, coding level, operational level, financial level
- Geographical location
  E.g. Natural shelter, vulnerability to specific natural disasters, terror targets, government
- Technical
  E.g. Media, hardware platform, hardware supplier, operating system, software packages for services

Bit Integrity - Relations between Elements

Which solution would you choose, if the costs were the same:

Solution 1: **number of copies element**	**Solution 2:** **number of copies & location independence elements**	**Solution 3:** **all elements**
100 copies	10 copies	3 copies
All placed at same active volcano	Placed at different locations with the same manager	Placed at different locations with different managers
No integrity checks	Only local integrity checks	Local & cross copies integrity checks

Risk management

- Risk can be thought of as an event or circumstance and the related uncertainties that may influence (e.g., prevent, impede or stop) the organization's achievement of a goal or affect business operations.
- Risk management is an iterative process.
- An organization's risk assessment will inform their use of the preservation storage criteria.

Cost Considerations

Things to consider when estimating and analyzing costs for digital preservation storage:

- Goals
- Scope
- Assumptions and dependencies
- Data collection
- Model your costs
- How do changes affect your costs?
- How do risks and uncertainty affect your costs?
- Document it, and continue to update based on actual data

Next steps

- Gather feedback on Usage Guide
- Incorporate feedback into Version 4
- Include cross references to applicable standards
- Investigate potential long-term host for the Criteria

More information

- Public Site at Open Science Framework (OSF): https://osf.io/sjc6u/
- dpstorage Google group: https://groups.google.com/forum/#!forum/dpstorage

Contributors

Contributors to version 1:

Goethals, A., Knight, S., Mandelbaum, J., Zwaard, K., McGovern, N., Truman, G., & Schaefer, S. (2016) What is Preservation Storage?. Workshop held at the Thirteenth International Conference on Digital Preservation, Bern, Switzerland. Abstract retrieved from https://phaidra.univie.ac.at/detail_object/o:502812

Additional contributors to current versions:

Zierau,E and Wu, C.