Preservation Storage Criteria, Version 2

Last modified May 2017

Introduction

Preservation Storage "provides the services and functions for the storage, maintenance and retrieval of [Archival Information Packages, or] AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which Archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders." [OAIS[1], 4-2] In this document, the term "Preservation Storage" is used, instead of Archival Storage or Digital Preservation Storage, but is understood to mean the storage for digital material under preservation management.

The Preservation Storage Criteria is a work in progress, intended to list the key criteria for preservation storage. Like the NDSA Levels of Preservation[2] it is intended to be an easy-to-use document that could help institutions just starting out with preservation as well as institutions with established preservation programs. Some of the uses for the Criteria include:

- Evaluating and comparing preservation storage solutions
- Determining gap areas in existing preservation storage
- Informing more detailed requirements for preservation storage
- As a component of instructional material on digital preservation
- To seed discussions with IT about preservation storage
- To seed discussions within the digital preservation field on preservation storage

The Criteria are meant to describe the characteristics of preservation storage that are relevant to a wide range of different kinds and sizes of institutions with responsibility for preserving digital material, or to organizations providing some or all of the preservation storage service to other institutions. The Criteria intentionally omits text that assumes specific architecture, technology, media, content, policy or vendor choices. It is not intended to be detailed enough to use as an institution's only preservation storage requirements document, and some of the criteria may not be applicable for all institutions. The Criteria are meant to be used by institutions as a foundation for informing their preservation storage, to be combined with local policies, applicable regulations, needs and preferences. Also note that institutions will need additional infrastructure in combination with preservation storage, e.g. staging areas, testing infrastructure, delivery and management servers.

Depending on the role an institution plays with regard to preservation storage, they may interpret the Criteria from a different perspective. For example, the "Documented access" criterion is defined as "Provides immutable logs and/or reports that show all file system access". A storage service provider could interpret this criterion to mean that they are responsible for providing the logs and reports; while an institution purchasing preservation storage from a vendor could interpret this criterion to mean that they expect to receive the logs and reports. Each of the Criteria could be interpreted as having a "providing" or "receiving" implication, depending on an institution's role in relation to the preservation storage.

This Criteria was originally developed by Kate Zwaard, Gail Truman, Sibyl Schaefer, Jane Mandelbaum, Nancy McGovern, Steve Knight and Andrea Goethals in preparation for an iPRES 2016 workshop[3] called "What is Preservation Storage?". It was also presented at the Designing Storage Architectures for Digital Collections 2016 meeting[4] hosted by the Library of Congress, and briefly at the PASIG Fall 2016 meeting[5]. Version 2 of the criteria reflects the feedback received at these meetings. A Google group has been set up to continue the preservation storage discussion and to solicit comments to keep refining the Criteria. To comment on this document or to join the conversation, please join the dpstorage group at https://groups.google.com/forum/#!forum/dpstorage.

Number	Criteria	Category	Description
1	Provides integrity checks	Content integrity	Performs verifiable and/or auditable integrity checking as part of the preservation storage
2	Supports independent integrity checks	Content integrity	Supports fixity checking by other parties, for example the content-owning institution
3	Provides preservation actions	Content integrity	Provides tools and/or services to support digital preservation actions (e.g. fixity checking, migration, auditing processes) as part of the preservation storage
4	Cost-efficient	Cost considerations	Costs relatively less than other more expensive solutions per GB, by being designed with cost efficiencies, for example, has resource pooling and sharing, multi-tenancy (multiple users share the same applications)
5	Energy-efficient	Cost considerations	Designed to conserve energy, for example, requires less cooling, consumes less power, uses less rack space, as in green computing initiatives
6	Storage weight	Cost considerations	The physical weight of the storage should meet certain qualifications, for example, be under a certain amount required for a particular floor.

7	High resilience	Flexibility & resilience	Has high resilience, which is the ability to adapt under stress or faults (e.g. resilient to equipment failures, power outages, attacks, surges in user demand)
8	High availability	Flexibility & 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
9	Recovery	Flexibility & resilience	Has documented ability to replace any corrupt/bad file, file system, or large-scale set of files in reasonable/expected/negotiated timeframes
10	Designed for zero data loss	Flexibility & resilience	Error detection and correction 24/7/365 (e.g. using RAID, Erasure coding, ZFS, triple copies/rebuild)
11	Independence	Flexibility & resilience	Storage layer is independent from other systems in the digital preservation environment so that it could be separately replaced without affecting the entire infrastructure
12	Nondisruptive storage migrations	Flexibility & resilience	Allows for storage tier changes over time (without disruption to availability)
13	Integratable	Flexibility & resilience	Storage layer is easily integrated with other systems and applications (i.e. plug and play)
14	Open source	Flexibility & resilience	Storage infrastructure can be integrated with open source tools and services in accordance with the organization's preferences
15	Supports multiple file system protocols	Flexibility & resilience	Infrastructure supports multiple file system protocols, e.g. NFS, CIFS, iSCSI, open or standard APIs etc. enabling vendor-neutral, direct addressability.
16	Diverse storage media types	Flexibility & resilience	Uses different storage media types together (e.g. disk and tape)
17	Management across multiple storage availability levels	Flexibility & resilience	Supports management and monitoring across multiple storage availability levels, e.g. online, near-line, off-line
18	Quality of storage media components	Flexibility & resilience	The known failure rate and technical characteristics of the storage media components is acceptable.

19	Secure	Information security	Includes safeguards, data security and documented procedures to prevent security incidents related to hardware, software, personnel, and physical structures, areas and devices.
20	Access controls	Information security	Provides role-based, access controls for storage infrastructure, e.g. user, staff, admin, to ensure only the appropriate people have the appropriate levels of access
21	Integration with authentication	Information security	Able to integrate with relevant organisational authentication systems to authenticate internal and external users of the system.
22	Deletion	Information security	Supports 'true deletion' (i.e. not just pointers) by authorised user or in accordance with local rules
23	At-rest server-side encryption with managed keys	Information security	Provides encryption at the storage layer, with no keys for customers to manage
24	At-rest server-side encryption with self-managing keys	Information security	Provides encryption at the storage layer, but customers manage encryption keys
25	Encrypted transfer	Information security	An appropriate transport layer encryption is used at all times when moving content
26	Multi-tenancy	Information security	Storage infrastructure supports separate roles/rules/access controls for separate agencies/departments/colleges/faculties etc
27	Virus/malware detection	Information security	Includes software that regularly runs virus checks and malware detection.
28	Virus/malware remediation	Information security	Provides remediation actions for content with viruses and/or malware, e.g. quarantine, notification, etc.
29	System error reporting	Information security	Provides immutable logs and/or reports that show all system errors, failures and other critical system activities
30	Supports expansion	Scalability & performance	Can increase storage over time as needed
31	Supports reduction	Scalability & performance	Can decrease storage over time to support deaccessions, transfer of ownership, etc.

32	Supports a global namespace	Scalability & performance	Nothing limits ability to have a global (i.e. consolidated) view of files
33	Use of multiple storage availability levels	Scalability & performance	Supports use of multiple storage availability levels, e.g. online, near-line, off-line
34	Tiered performance	Scalability & performance	Meets specified/negotiated performance levels appropriate to material being stored, e.g. Tier1 storage for metadata indexing and searching, Tier2 for caching, Tier3 or lower for bulk storage.
35	Scalable to large data sizes	Scalability & performance	Able to support very large amounts of content, e.g. multiple PBs of data, hundreds of millions of files and directories, terabyte size files
36	File system limits	Scalability & performance	Able to support long file, path or directory names; large amount of files in a directory, diverse character encodings
37	Delivery	Scalability & performance	Meets expectations for delivery from the storage layer, e.g. at a reasonable/negotiated rate and supporting concurrent users
38	Complete exports	Scalability & performance	Supports the bulk exporting of content and metadata for any reason, at an acceptable rate, for example, as part of an exit strategy
39	I/O performance	Scalability & performance	The input/output performance of the system or service is at an acceptable rate
40	Compute power	Scalability & performance	Computing power of the system or service is at an acceptable rate and available when needed
41	Geographic separation	Storage location	Ensures multiple redundant copies in geographically-separate locations for protection from catastrophic loss
42	Replication	Storage location	Has documented ability to create redundant, distributed copies of content in reasonable timeframes
43	Customizable replication based on content	Storage location	Storage infrastructure supports content-specific user-defined replication rules, for example less copies of a particular stream of content
44	Expose and constrain location	Storage location	Storage infrastructure exposes the specific storage location of data to meet content-specific requirements (e.g. location constraints, transparency of expectations and requirements)

45	Support commitment	Support	Documented vendor or IT commitment to support storage infrastructure, e.g. through SLAs (addressing for example responsibilities, data assurance, response times, end-of-service exit provisions, etc.)
46	Training	Support	Training provided to appropriate staff across all relevant operational and maintenance tasks
47	Accessibility	Support	Ensures people with disabilities equivalent access to reports, documentation and other content
48	Supports open storage formats	Transparency	Infrastructure supports open, standard, non-proprietary storage formats, e.g. TAR, archive eXchange format (AXF), LTFS
49	Data error notification	Transparency	Notifies content-owners of all data errors, remediation actions and issues in reasonable/expected/negotiated timeframes
50	Self-healing transparency	Transparency	Systems that use mechanisms to correct altered data (like bit corruptions) do so in a transparent, documented manner.
51	Supports independent preservation actions	Transparency	Supports digital preservation actions (e.g. migration, auditing processes) by other parties or external tools, for example a format migration by the content-owning institution running tools that are not part of the storage infrastructure
52	Monitoring	Transparency	Supports ability to observe or check activity in the storage infrastructure (e.g. see activity in real-time, examine logs, observe the performance status, determine the overall status or drill-down into activities)
53	Provides content reports	Transparency	Provides reports about content in the storage infrastructure (e.g. number of objects/files/formats, average file size, types of objects, size of storage in use)
54	Provides activity reports	Transparency	Provides reports about activity in the storage infrastructure (e.g. fixity or virus results, corruption, replacement with good copies)
55	Custom reports	Transparency	Supports custom (for example configurable and/or on-demand) reporting of content or activity in the storage infrastructure
56	Documented infrastructure	Transparency	Provides full, complete, current, and available documentation of key processes, services, systems, procedures, known limitations and functions
57	Documented access	Transparency	Provides immutable logs and/or reports that show all file system access

58	Documented provenance	Transparency	Documents audit/provenance information about all changes, for example about integrity check failures, deletions, modifications, additions, preservation actions; and who or what performed the actions

Footnotes

1. Consultative Committee for Space Data Systems. (2012). Reference Model for an Open Archival Information System (OAIS), Recommended Practice, CCSDS 650.0-M-2 (Magenta Book) Issue 2. <http://public.ccsds.org/publications/archive/650x0m2.pdf>
2. Phillips, M., Bailey, J., Goethals, A., & Owens, T. (2013). The NDSA Levels of Digital Preservation: An Explanation and Uses. <http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf>
3. Goethals, A., Knight, S., Mandelbaum, J., Zwaard, K., McGovern, N., Truman, G., & Schaefer, S. (2016) What is Preservation Storage?, Workshop at iPRES 2016.
4. Mandelbaum, J. (2016). Preservation Storage Criteria, Presentation at Designing Storage Architectures for Digital Collections 2016, Library of Congress.
5. Schaefer, S. (2016). Preservation Storage Criteria, Lightning talk at PASIG Fall 2016 Meeting.