An Overview of the Digital Preservation Storage Criteria and Usage Guide

Eld Zierau
Royal Danish Library, Denmark
elzi@kb.dk
0000-0003-3406-3555

Nancy Y McGovern
Massachusetts Institute of Technology, USA
nancymcg@mit.edu
0000-0002-7733-1516

Sibyl Schaefer
University of California, San Diego, USA
sschaefer@ucsd.edu
0000-0002-7292-9287

Andrea Goethals
National Library of New Zealand, New Zealand
Andrea.Goethals@dia.govt.nz
0000-0002-5254-9818

Abstract – The Digital Preservation Storage Criteria (or "Criteria") resulted from a community discussion at iPres 2015 on providing guidance to organizations that either use or provide digital preservation storage. First developed in 2016, they have been refined in iterative versions over the last three years based on feedback gathered at conference sessions and through a survey. The Criteria are intended to help with developing requirements for, or evaluations of, preservation storage solutions; to seed discussions about preservation storage; or to use within digital preservation instructional material. The latest version of the Criteria contains sixty-one criteria grouped into eight categories: content integrity, cost considerations, flexibility, information security, resilience, scalability & performance, support, and transparency.

The key new development since the Criteria was presented at the iPRES 2018 workshop is a usage guide, developed to accompany the Criteria. It includes sections on key topics to consider for preservation storage in addition to the Criteria: risk management, independence, elements in establishing bit safety, and cost considerations. The usage guide will be released publicly for review as one of the next steps in the project, along with developing version 4 of the Criteria and taking steps to further build the community around the Criteria.

Keywords – digital preservation storage, archival storage, criteria, risk management

Conference Topics – Designing and delivering sustainable digital preservation; The cutting edge: technical infrastructure and implementation; Collaboration: a Necessity, an Opportunity or a Luxury?

I. Introduction and Background

The Digital Preservation Storage Criteria (or "Criteria") are a result of a collaborative process based within the digital preservation community. This paper provides some context that traces the development and implementation of the Criteria and looks ahead to current and possible future developments. The development of the Criteria has involved iterative cycles of definition and elaboration by a working group, followed by opportunities for community review and feedback, and then finally the integration of community feedback into a series of versions that are publicly available on a project website [1]. Since the advent of computers, storage and processing capacity have framed the development and evolution of preservation strategies; the Criteria are meant to address evolving organizational requirements as digital preservation programs mature, as technological options emerge and evolve, and as opportunities and challenges become clearer.

A. Definition of Digital Preservation Storage

One of the prerequisites for identifying and elaborating the Criteria was developing a working definition of Preservation Storage, absent a shared and authoritative definition within the digital preservation

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

community. Defining "digital preservation storage" requires first defining "digital preservation." The definition adopted as a starting point is from the Digital Preservation Coalition: "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary" [2].

Building on this base definition, the working definition of digital preservation storage for the purposes of the Criteria is: a fundamental component of digital preservation that supports and enables ongoing digital preservation activities. The term digital preservation storage encompasses the functions of the OAIS [3] functional entity *Archival Storage* as well as related OAIS functional entities that are needed to store, maintain in storage, and retrieve Archival Information Packages (AIPs) from storage [4].

For example, preservation storage includes parts of the following:

- ***Preservation Planning*** responsible for monitoring technology for storage options, relevant standards and practices, and media migrations.
- ***Data Management*** that ensures the relationship between preserved content and its associated metadata.
- ***Administration*** concerned with policies and standards pertaining to preservation storage management.
- ***Ingest*** concerned with the coordination of input and updates to different data replicas in storage.

The Criteria are intended to continually enable the digital preservation community to weigh the potential opportunities and risks of modern storage services and options while addressing the expectations of modern digital preservation practices.

B. Background on the Criteria Creation

The roots of the Criteria trace back to an initial digital preservation community discussion of digital preservation storage that was convened by the iPres 2015 conference organizers, which in part highlighted the lack of a guiding document related to preservation storage. Several of the participants then put forward a call for volunteers to establish a working group to design a set of preservation storage requirements. It quickly became clear that "requirements" would vary from organization to organization, and thus were unrealistic and unhelpful to outline. What was helpful was a list of criteria from which to select and further develop into specific requirements. Thus, the Criteria were born.

The working group culled requirements from several Requests for Proposals that they had used in various organizational settings, and then abstracted specific requirements into more general criteria. In preparation for the 2016 iPres workshop on the Criteria, the working group listed these starter criteria in a survey that was delivered to workshop participants prior to the conference. The survey asked participants to rank each criterion according to their value. This activity was successful in getting the participants to engage deeply with the Criteria and the result was a productive conversation during the workshop. The feedback generated in this iPres workshop, as well as during an earlier workshop held at the annual Library of Congress Designing Storage Architectures meeting, was then incorporated into the second version of the Criteria.

The Criteria working group then used this same pattern -- revision of the Criteria, presentation and workshopping of them at iPres and the Library of Congress Designing Storage Architectures meetings, followed by incorporating feedback to create a new version -- during 2017 and 2018. The working group also created a Google email group[1] for interested community members to discuss and comment on the work and new versions.

The working group is currently drafting version 4 following a series of presentations at 2018 conferences and a workshop at iPres 2018[2].

C. Potential Uses and Audiences

The Criteria have been developed as a set of design attributes, and considerations for digital preservation storage services. Some of the uses for the Criteria include:

[1] See groups.google.com/forum/#!forum/dpstorage

[2] The forums where the Criteria has been presented for community feedback are listed on the project website wiki (osf.io/sjc6u/wiki).

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

- Guiding evaluations of preservation storage services and options
- Identifying gaps in existing digital preservation storage implementations
- Assisting with Request for Proposals (RFPs) and related documents
- Contributing to instructional materials on digital preservation
- Informing infrastructure design and planning with Information Technology (IT) and other domains
- Framing discussions within the digital preservation community.

The possible audience(s) for the Criteria include digital preservation managers who need to implement and manage digital preservation storage, providers of digital preservation storage services, auditors of digital preservation programs, digital preservation instructors and students, and practitioners in affiliated domains who rely upon digital preservation storage.

A guiding principle for the versions of the Criteria has been ensuring that the Criteria remain generally applicable to digital preservation storage in any context by avoiding the inclusion of local practices. The Criteria provide a bridge to implementation by including a usage guide and accumulating examples to demonstrate the local use of the Criteria.

II. Structure of the Criteria

A. *Presentation*

The Criteria are organized into a table with five columns and one row per criterion shown in Table 1.

TABLE I

Structure of the Preservation Storage Criteria

No.	Criteria	Category	Description	Related Criteria & Refe-rences
1	Integrity checking	Content Integrity	Performs verifiable and/or auditable checks to detect changes or loss in or across copies ...	
2	...	...	...	
...				
61	...	...	...	

The columns are for the 'Number' (sequential ID for the criterion), 'Criteria' (short descriptive name for the criterion), 'Category' (one of eight topical areas used to group the Criteria), 'Description' (short definition for the criterion), and 'Related Criteria and References' (a placeholder to map relevant standards or related criteria to the criterion). For example, in Table 1, the first listed criterion is "Integrity Checking" in the category of "Content Integrity." The Integrity Checking criterion indicates that the preservation storage "Performs verifiable and/or auditable checks to detect changes or loss in or across copies." There currently are no related criteria or references listed for this criterion.

B. *Categories*

Starting with the second version of the Criteria, the initially unwieldy list of criteria was organized into categories to group similar criteria together and to provide an overall organization. Currently, the eight categories are:

1. **Content Integrity** refers to practices ensuring the state of stored data has not changed.
2. **Cost Considerations** reflect the financial impact of storage decision making.
3. **Flexibility** refers to the adaptability, interoperability, and overall ability to customize preservation storage solutions to an organization's needs.
4. **Information Security** refers to data protection methods to ensure that the data cannot easily be tampered with or accessed without proper authorization.
5. **Resilience** refers to the durability and availability of the storage system.
6. **Scalability & Performance** refers to computational performance and ability to be scaled up or down according to organizational needs.
7. **Support** refers to support contracts as well as services like training and additional preservation services such as migration.
8. **Transparency** refers to the visibility into the storage system's functions, e.g. auditing, reporting, error notification, and documentation.

C. *Revisions*

As mentioned previously, the Criteria have been revised several times because of feedback from workshops, presentations at conferences, and a survey. The introductory narrative of the current version of

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

the Criteria (version 3) has been enhanced to add:

- more clarity on the definition and scope of "preservation storage"
- clarification that the audience for the Criteria includes both consumers and providers of preservation storage
- additional key considerations to consider in addition to the Criteria

Changes were also made to the Criteria table to include categories (see Table 2) and to normalize the Criteria names (bolded) and definitions. Finally, a reference list and an accompanying usage guide were developed.

TABLE II

Evolution of the Criteria Categories

	2016 - Version 1	2017 - Version 2	2018 - Version 3
No. of Criteria	48	58	61
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)

IX. Usage Guide

The Criteria cannot stand alone; they need to be set in context of basic preservation principles. Therefore, the Criteria are supplied with a usage guide focusing on preservation storage principles.

Preservation is about preventing the loss of data, therefore managing the risks that could cause data loss is an essential practice for all types of preservation. The usage guide therefore includes the following key concepts that should be considered in relation to the Criteria: risk management, **independence between copies, elements in establishing bit safety and cost analysis.**

A. Risk Management

The usage guide includes a short description of the general concepts and processes of the practice of risk management to help organizations using the Criteria. Digital preservation requires storage solutions that can be sustained over the long-term. Risks to digital preservation storage operations may come from one or many events, incidents or situations. The usage guide includes a list of examples of these.

An organization can use risk management practices to identify and isolate risks that are specific to digital preservation over the long-term to reduce and mitigate impacts on digital preservation operations. Similarly, an organization can use a risk assessment to compare the risks of storage solutions that address different sets of criteria. Because digital preservation storage solutions must be sustained over time, it is useful to have a consistent methodology for risk management that can be used by the organization over time as solutions change, and as organizations use the Criteria to propose solution changes over time.

B. Independence Between Copies

For Preservation Storage, risk management must consider the goal that no or only an acceptable amount of data may become lost. There are risks that one event, agent, or technology can harm several copies of data in a way which imply loss of all data or an unacceptable possibility of data loss. The best way to mitigate such risks is to ensure independence between copies in a way that prevents the same event or incident for doing such harm. Independence means that any one event, agent, or technology cannot affect a majority of copies. The independence must be considered on any level where risk of loss can exist, e.g. organizational level, technical level, environment level etc. The total risk assessment must take all three key elements (number of copies, independence between copies, and integrity checks of and among copies) into account for each type of risk.

It is important to note that independence between copies may include the use of checksums. This is especially the case when there is a minimum number of copies (two full copies and one checksum), since loss of both checksum and one copy will make it impossible to verify whether the surviving copy is correct.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

C. *Elements in Establishing Bit Safety*

A full risk assessment of Preservation Storage needs to include more than independence between copies; it needs to include all the three essential elements which are needed for evaluating whether a Preservation Storage solution provides the required level of bit safety. These are:

- **Number of copies** - There should be enough copies available to survive the loss of some number of the copies.
- **Independence between copies** - The copies should exist independently of one another
- **Integrity checks (of copies and among copies)** - The copies must undergo periodic integrity checks to assure their fidelity.

The decisions on how many copies are needed can be determined with a complete risk assessment with focus on risk of losing all copies or losing the ability to verify correctness of surviving copies based on consideration of all three elements. Risk assessment may vary due to which risks each organization is willing to take. The absolute minimum number of copies is two, since an error in one copy requires having a healthy copy to be repaired from. The risk of keeping only two copies is that unless information like a checksum is also kept, there may be no way to tell which copy is valid if one becomes erroneous. When using such a minimum setup it is very important to consider the risks of loss.

Another important part of Preservation Storage is to consider how requirements for confidentiality and availability and costs of the preserved data are ensured, e.g. it may be hard to ensure confidentiality for data that has 100 copies spread all over the world, and it may be difficult to provide fast access to data that is only placed on off-line media. Such issues need to be considered as part of the risk analysis along with the other bit preservation elements.

D. *Cost Analysis*

The usage guide includes a short description of the general concepts and processes of the practice of cost analysis, to help organizations using the Criteria. An organization can use cost analysis to identify and isolate storage solution costs that are specific to digital preservation, and/or to compare the costs of different storage solutions that address different sets of criteria.

Cost analysis is a systematic approach to estimating resource expenditures, either to compare potential or existing situations, or to establish an approach for valuing resources for a specific decision or course of action. For example, a cost analysis can help identify and compare the resources required to implement and sustain two different storage solutions which are based on different sets of digital storage criteria. The usage guide includes an introduction to cost assessment and how it is used as well as tools and additional resources.

X. Future Work

While much of the content of the usage guide was presented in recent iPRES and PASIG conference sessions, at the time of this writing the usage guide has not been released. The next step for the working group is to complete the first version of the usage guide and to release it publicly for feedback by members of the dpstorage Google group and the broader digital preservation community.

There is also work planned for the Criteria document itself. Version 4 of the Criteria will map the Criteria to applicable standards and will incorporate feedback from recent conferences.

Additionally, effort will go into building the community around the Criteria project. The project website will be improved to expose more of the project outputs and roadmap. Examples of organizations using the Criteria will be documented and shared through the project website. Lastly, an organizational host for the Criteria project will be sought to provide a stable home for the Criteria and to help engage the community to use and improve it.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

Acknowledgment

The latest version of the Criteria has been shaped by the feedback of many individuals in the digital preservation community over the last three years. In addition to the authors of this paper, core members of the Digital Preservation Storage Working group include Jane Mandelbaum, Gail Truman, and Cynthia Wu. Steve Knight and Kate Zwaard made important contributions to early versions. People who joined the dpstorage Google Group have provided valuable feedback. Special thanks to the attendees of the iPRES 2017 Digital Preservation Storage workshop in Kyoto who gave us in-depth feedback on the Criteria and the newly introduced categories.

References

[1] A. Goethals, N. McGovern, S. Schaefer, G. Truman, and E. Zierau. 2018. Digital Preservation Storage Criteria. Retrieved June 19, 2019 from https://osf.io/sjc6u/ DOI: 10.17605/OSF.IO/SJC6U

[2] Digital Preservation Coalition. 2015. Digital Preservation Handbook (2nd. ed.). Retrieved March 13, 2019 from https://dpconline.org/handbook

[3] ISO 14721:2012. "Space data and information transfer systems - Open archival information system (OAIS) – Reference model", 2012.

[4] N. McGovern and E. Zierau. "Supporting Analysis and Audit of Collaborative OAIS's by use of an Outer OAIS - Inner OAIS (OO-IO) Model," Proceedings of the 11th International Conference on Preservation of Digital Objects (iPres 2014), pp. 209-218, 2014.

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.