

This document was published as:

Rethinking digital preservation: definitions, models, and requirements

Thomas C. Wilson

Digital Library Perspectives Volume 33, Issue 2 pp. 128-136

<https://doi.org/10.1108/DLP-08-2016-0029>

Rethinking Digital Preservation: Definitions, Models, and Requirements

As digital preservation activities have matured and the options for tools and services have increased, it is worth revisiting what is meant by digital preservation in order to differentiate the marketplace of options. It is fair to say that there are many shades of gray in this space and that clarity on definitions and models will serve to ensure the community at large achieves the outcomes it seeks. This article will focus on both conceptual and pragmatic issues in digital preservation. The timing for such a review is critical as the digital library and digital preservation communities begin a review of the foundational, but now dated, OAIS Reference Model. (Rosenthal, 2015; Digital Preservation Coalition, 2016b) It is also helpful to view the relationship between OAIS and related audit and certification standards, e.g., TRAC, ISO 16363:2012.

What is Digital Preservation Now?

Digital preservation now includes multiple options to consider from commercial, non-profit, consortial, or open source providers. In order to understand fully the applicability of any given solution, one must grasp a number of fundamental issues underlying digital preservation. More to the point, for audit and certification processes it is essential to be explicit about the relative strengths of specific solutions and strategies to the challenges of digital preservation.

It is worth detailing what constitutes digital preservation and what does not. One can debate a variety of issues involved in digital preservation, but a basic premise is that digital preservation is not just one thing; it is a cluster of many practices, policies, technologies, and structures. (Digital Preservation Coalition, 2016a; Corrado and Moulaison, 2014, p. 3-7; Hedstrom, 1998) On a fundamental level it is also worthwhile to outline things that are necessary, but not sufficient, prior to embarking on a preservation program. An obvious requirement is to have a repository of digital objects of sufficient perceived value to warrant an investment in a preservation mechanism. Another is a tested backup and recovery system as part of a larger disaster recovery procedure. Although these items are necessary, they do not constitute digital preservation in and of themselves.

A 2011 Association of Research Libraries SPEC Kit survey on digital preservation highlights another differentiation of the meaning of digital preservation. The responses from the participants indicate that digital preservation can mean:

- The preservation of rare and/or brittle materials through digitization, a mechanism for reducing the direct handling or manipulation of the materials; or
- The preservation of licensed digital content, e.g., journal articles; or

- The preservation of locally digitized content, typically materials from special collections or archives; or
- The preservation of born digital content, such as institutional records. (McMillan et al., 2011)

From a broad perspective all of these activities produce digital materials that require preservation of the digital form. Treating each of these as a different means of digital preservation, however, creates confusion over what is being preserved in the process of digital preservation from the perspective of the OAIS model. The responses to the survey also indicate a conflation of parts of and the whole of digital preservation, e.g., interpreting digital preservation planning activities, such as, articulating policies and procedures or collecting metadata as participating wholly in digital preservation. While advocating for any of the numerous activities involved in the digital preservation process is recommended, it is likewise necessary to understand if and when some material is actually preserved “digitally.” Rothenberg summarizes this point, “The preservation and management of digital records involves interrelated technical, administrative, procedural, organizational, and policy issues, but a sound technical approach must form the foundation on which everything else rests.” (1999, p. 6)

Holdsworth adds into the mix the notion, “The key to doing things properly is to take a view of digital data as an abstract quantity, divorced from the medium upon which they are stored, but associated with information (technical metadata – often including software) that permits ready access to its intellectual content.” (2006, p. 34) Noting the impossibility of preserving the ability to read the original medium, he also states, “The approach is to preserve this byte-stream indefinitely, copying it unchanged as storage technology evolves.” (2006, p. 41) Holdsworth uses layers of abstraction to support emulation, “access to intellectual content” at the time of use, and media independent data. (2006, p. 42-44)

It is useful to understand that digital preservation is a game of probabilities. The activities undertaken are done to reduce the likelihood that a given object will be lost or corrupted because of bit rot, nefarious actions, obsolescence, etc. There is no 100% guarantee that all things will be saved and preserved; objects have also been lost in the physical world. The sheer scope and range of digital objects strongly suggest the need for selection of what is to be preserved. (Kastellec, 2012) The goal is to perform certain sets of actions that together mitigate the risks associated with digital objects. Within these sets of actions, there are options, as well as levels of commitment that enhance the robustness of a given digital preservation strategy and implementation. In other words, there is a hierarchy of digital preservation needs.

Abraham Maslow developed a theory of human motivation and development that is expressed in the form of a pyramid indicating a hierarchy of needs beginning at the bottom with basic physiological needs and ending at the pinnacle with self-actualization, each step up the pyramid requiring the lower needs to have been met. (Maslow, 1954) Although Maslow’s specific conclusions have been called into question with regard to human motivation and development, the modeled layering of priorities and dependencies serves as a useful tool for other topics. A

similar approach can be applied to digital preservation by starting with the foundation of a collection of digital objects and associated metadata in a repository and proceeding up the hierarchy to a fully auditable digital preservation mechanism.

The value of this model can be found in its adherence to an understanding that digital preservation is not just any one or several of these activities, it is all of them in combination building to a robust preservation system. The model also conveys a priority order of activities in a way that suggests a pathway for beginning a digital preservation process, as well as providing a practical mechanism for identifying areas not addressed by a particular existing solution or process. It functions at a lower level of abstraction than the OAIS Reference Model; it does not replace it. Some members of the preservation community emphasize metadata creation and policies and procedures as foundational elements of a digital preservation system. Although these two areas are critically important in the overall scheme of digital preservation, their presence alone does not preserve anything digitally.

Hierarchy of Digital Preservation Needs

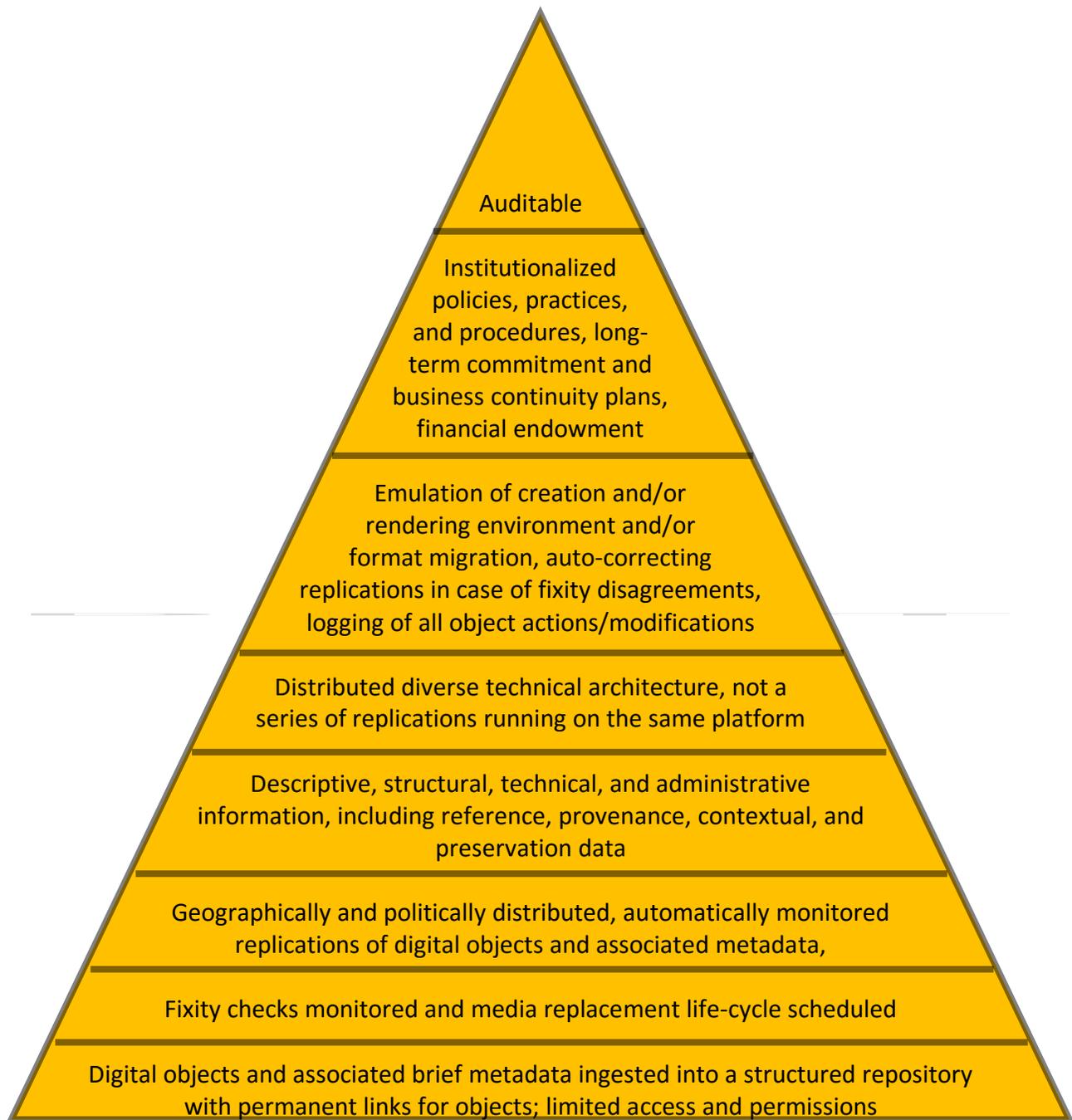


Figure 1 [1, 2]

While one can debate the relative placement of these activities within the hierarchy, two statements remain true:

- Bit level preservation and appropriate object metadata are both necessary. Without both, a digital preservation system does not exist.
- Adding more of these activities creates a more robust system. A corollary of this statement is: something is better than nothing. (Digital Preservation Coalition, 2016c)

When planning and implementing a digital preservation solution, it is essential to identify points of weakness in the overall system(s). Thus, in the hierarchy having monitored distributed replications is better than having a single instantiation or distributed, but unmonitored, replications. And having replications distributed across multiple platforms is even better. The point here is to reduce or eliminate single points of failure whether hardware, operating system, and/or software.

This would be a good point at which to emphasize the absence of the word "cloud" in this discussion. Although "the cloud" can imply a certain type of technical infrastructure, it has become mostly a marketing term that can obfuscate what is really happening. For example, just because something is stored in "the cloud" does not mean that it is replicated or monitored. Without those two functions, a digital preservation system does not exist, whether in "the cloud" or elsewhere.

OAIS Reference Model

The OAIS Reference Model has long served as a framework for understanding the components and relationships involved in digital archiving. (Lee, 2010; Sawyer et al., 2002) It has functioned as a guide for conceptualizing and developing systems to address preservation needs at an abstract level. Reference models can be helpful in representing required functional components of a class of complex systems designed to address large scale shared needs. They also support a modular approach to building the components of a system(s). "The reference model does not specify a design or an implementation. Actual implementations may group or break out functionality differently." (The Consultative Committee for Space Data Systems, 2012) (CCSDS)

Several challenges arise when a reference model, or its framework, is taken too literally in system design. Some particularly thorny pitfalls are:

- Singular system vs. multiple systems
- "System" in a more global sense
- Preservation vs. discovery and delivery

It is relatively easy to conclude, when reviewing the OAIS Reference Model, that it requires a single system to provide all of the component functions in the model. This is not the case, however. If this were true, the model would violate one of the foundational principles of digital preservation—avoid single points of failure. Indeed, some commercial systems fall into this trap.

In today's environment of digital objects, it is also convenient to assume that the "system" is a digital one. This assumption would be unfortunate as the model itself includes decidedly human components (e.g., policy determination). It is appropriate to see the model as applicable to completely non-digital archiving, as well as having many and various components that contribute to an overall archival and preservation set of systems.

As an outgrowth of the first two challenges, the dual roles of a preservation system, as represented in the model, preservation and access, can become needlessly technically intertwined. While the reason for preservation may be for future access, mixing philosophical issues and technical architectures can lead to confusing outcomes and potentially compromise the integrity of a preservation archive. In practice often the Dissemination Information Package includes a digital object that is, in fact, different than the object in the Archival Information Package (e.g., resolution and file type for an image object). [3]

In addition, by separating the preservation component from the access component, it is possible to build a more robust preservation system by forming a dark archive with limited access for the purposes of long-term preservation. Furthermore, this separation also supports differentiated development pathways for access and preservation tools. One can argue that the OAIS Reference Model makes this separation clear; however, when the model is used literally as a guide for the purposes of certifying trustworthy repositories, separate systems or distributed storage systems may not be viewed as acceptable, i.e., trustworthy and certifiable.

The Digital Curation Centre's Curation Lifecycle Model is a more recent development that appears to clarify matters by emphasizing objects and actions rather than systems. It also includes an explicit understanding that "users of the model may enter at any stage of the lifecycle" and that different organizations may provide different pieces of the overall curation and preservation mechanisms. [4]

Seles (2016) outlines limitations of both OAIS and the process through which this standard and related ones are generated. She notes that OAIS is a reference standard, so there is no OAIS compliancy by definition; the use of the phrase in the literature commonly reflects this misunderstanding. (e.g., Houghton, 2015) Seles cites several concerns about the standards process in general and the OAIS and related standards experience in particular:

- Communities served by the standard in relation to committee membership.
- Lack of transparency in the process of committee membership.
- Lack of full engagement and consideration of the cultural heritage community, privileging other specific perspectives.
- Limited sample of types of archives in the testing phase.
- Focus on idealized environments that may not reflect the situations that archiving entities encounter.
- Notion that a lack of a written mandate for organizations to archive specific materials would automatically cause them to fail the audit and certification process.

- Lack of a consensus across the broad community using the OAIS Reference Model and inclusion of and sensitivity to differing points of view.

To address these concerns, Seles recommends removing the review process from under the CCSDS structure, balancing theory and practice in both the OAIS and the related audit and certification standards, pressuring the ISO to enforce technical committees' and subcommittees' engagement with liaison committee guidance, and increasing transparency and inclusion by the OAIS revision committee.

To see some of the interplay between interpretations of the existing OAIS Reference Model, one need to look no further than Giaretta's (2016) response to Rosenthal's (2015) posting on the need for OAIS revision or the Digital Curation Centre and Digital Preservation Coalition response to the CCSDS comments on recommendations at the five-year review of OAIS. (Higgins, S. And Boyle, F., 2008) While there is room for healthy debate on statements and their interpretation within the model, a more critical issue is how those statements and interpretations have been codified into audit and certification standards that have significant impact on which systems and methods receive approval. It is imperative that the review and updating of the OAIS model now underway explicitly clarify the appropriate interpretation and use of OAIS as it relates to these other standards.

Audit and Certification Challenges

While the OAIS Reference Model can be used to determine if the functional preservation requirements are met by any given system(s), audit and certification process checklists, such as "Trustworthy Repositories Audit & Certification: Criteria and Checklist", in addition focus on organizational issues such as permanency, structural integrity, commitment, staffing, resources, business continuity, etc. When the reference model is taken too literally, or interpreted at a lower abstraction level than intended, a solution for digital preservation that does not address dissemination to consumers, would be deemed in violation of the model and, therefore, non-compliant, i.e., untrustworthy.

Section B5. Information Management of the "Trustworthy Repositories Audit & Certification: Criteria and Checklist" states:

- "Regardless of system, descriptive information (metadata) will be acquired and maintained for access and retrieval. If people cannot find what they want, the repository is not serving the needs of its users."
- "People also need to know whether they are permitted to get a usable copy of it and how."
- "A repository's minimum descriptive metadata requirements must match the minimum needs of the repository's designated community(ies)." (OCLC & CRL, 2007, 35)

These statements, all within two paragraphs, belie an assumption of a live repository that a (general) public or identified community would use to access digital materials as a part of a trustworthy repository with a preservation component. In other words, a dark archive could not be certified as a trustworthy repository, when, in fact, it may actually achieve a more robust preservation strategy.

There are problems with these criteria and checklist statements:

- Users of such a dissemination system may need to know what permissions they have for the use of an object, but they do not necessarily need to request the object; they already have a version of it in front of them, a function of modern system design in the user interface.
- In the case of a dark archive, the user community could be as few as one. The use would be to recover an object to replace a damaged Dissemination Information Package in the delivery mechanism which is implicitly aimed at a larger audience.

All of this is to say that, when the OAIS Reference Model, as written, is used as a basis for certifying trustworthiness—a high stakes activity—and interpreted literally without recognition of the nature of a reference model, the outcomes are not only short-sighted, but may also lead to advancing system design that does not fully advocate for robust digital preservation.

Conclusion

In order to judge the nature, extent, and robustness of a digital preservation solution, one must view it against a holistic model that clearly indicates the relative and comprehensive merit of all components. The hierarchy of digital preservation needs presented here offers a model based on nearly twenty years of experience in real world digital preservation across multiple communities on gaging the strength of digital preservation options.

It is fitting that the OAIS Reference Model, and the process creating it, be reviewed and updated by the communities affected and that audit criteria, documentation, requirements, and checklists be revised to reflect current knowledge and experience broadly to support the development of demonstrably more robust digital preservation solutions. Furthermore, the revised OAIS Reference Model and related certification standards need to include explicit language that clearly reflects an understanding that a multi-system architecture is acceptable and that a dark archive design can be compliant. Absent this outcome, portions of the cultural heritage community will need to develop, modify, or crosswalk to a more flexible set of criteria. If the resultant collection of certified repositories includes only those hosted by well-resourced organizations, another set of problematic issues arises, most important of which is how does this community provide robust options that are achievable by the less well-resourced members of the cultural heritage community.

Notes:

1. These levels are partially inspired by the NDSA Levels of Digital Preservation [<http://ndsa.org/activities/levels-of-digital-preservation/>] (accessed 12 July 2016) with a bit more direction and a collapsed hierarchy.
2. Please note: there is no mention of backups. Backups do not qualify as digital preservation as they are not monitored. Also, no media that degrades with each reading (e.g., tape) qualifies as preservation storage.
3. Nancy McGovern wrote into the DPOE training material the importance of separating preservation and access. See DPOE Module #4 - Provide, slide 3 at http://projects.iq.harvard.edu/ndsr_boston/educational-enrichment (accessed 31 July 2016).
4. The DCC Curation Lifecycle Model is available at: <http://www.dcc.ac.uk/resources/curation-lifecycle-model> (accessed 31 July 2016).

References:

The Consultative Committee for Space Data Systems. (2012), "Reference model for an open archival information system (OAIS), magenta book", pp, 1-2, available at: <http://public.ccsds.org/publications/archive/650x0m2.pdf> (accessed 20 June 2016).

Corrado, E.M. and Moulaison, H.L. (2014), *Digital Preservation for Libraries, Archives, and Museums*, Rowman & Littlefield Publishers, Lanham, MD.

Digital Preservation Coalition. (2016a), "Digital preservation briefing", in DPC, *Digital Preservation Handbook*, Digital Preservation Coalition, Glasgow, UK, available at: <http://handbook.dpconline.org/digital-preservation> (accessed 14 July 2016).

Digital Preservation Coalition. (2016b), "OAIS community", available at: http://wiki.dpconline.org/index.php?title=OAIS_Community (accessed 14 July 2016).

Digital Preservation Coalition. (2016c), "Standards and best practice", in DPC, *Digital Preservation Handbook*, Digital Preservation Coalition, Glasgow, UK, available at: <http://handbook.dpconline.org/digital-preservation> (accessed 14 July 2016).

Giaretta, D. (2016), "Comments on David Rosenthal's 'The case for a revision of OAIS'", available at: http://wiki.dpconline.org/index.php?title=Comments_on_David_Rosenthal%27s_%E2%80%9CThe_case_for_a_revision_of_OAIS%E2%80%9D (accessed 14 July 2016).

Hedstrom, M. (1998), "Digital preservation: a time bomb for digital libraries", *Computers and the Humanities*, Vol. 31, pp. 189-202.

Higgins, S. and Boyle, F. (2008), "Response to CCSDS's comments on the OAIS five-year review: recommendations for update, 2006", available at: http://www.dpconline.org/component/docman/doc_download/133-response-to-ccsds-comments-on-the-oais-5-year-review-recommendations-for-update-report-response-to-ccsds-comments-on-the-oais-5-year-review-recommendations-for-update-report?q=response+ccsds%27+comments (accessed 25 August 2016).

Holdsworth, D. (2006), "Strategies for digital preservation", in Deegan, M., Tanner, S. (Eds.), *Digital Preservation*, Facet Publishing, London, UK.

Houghton, B. (2015), "Trustworthiness: self-assessment of an institutional repository against ISO 16363-2012", *D-Lib Magazine*, Vol. 21 No. 3/4, available at: DOI: 10.1045/march2015-houghton (accessed 25 August 2016).

Kastellec, M. (2012), "Practical limits to the scope of digital preservation", *Information Technology and Libraries*, Vol. 31 No. 2, pp. 63-71, available at: <http://dx.doi.org/10.6017/ital.v31i2.2167> (accessed 1 June 2016).

Lee, C.A. (2010), "Open archival information system (OAIS) reference model", in Bates, M. (Ed.), *Encyclopedia of Library and Information Sciences, Third Edition*, Taylor & Francis, Oxfordshire, UK, pp. 4020-4030, available at: <http://ils.unc.edu/callee/p4020-lee.pdf> (accessed 9 May 2016).

Maslow, A. (1954). *Motivation and personality*, Harper, New York, NY.

McMillan G., Schultz, M., & Skinner, K. (2011), *SPEC Kit 325: Digital Preservation*, Association of Research Libraries, Washington, DC.

OCLC & CRL (2007), "Trustworthy repositories audit & certification: criteria and checklist", p 35, available at: https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf (accessed 20 June 2016).

Rosenthal, D. (2015), "The case for a revision of OAIS", available at: http://wiki.dpconline.org/index.php?title=The_case_for_a_revision_of_OAIS (accessed 14 July 2016).

Rothenberg, J. (1999), *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*, Council on Library and Information Resources, Washington, DC.

Sawyer, D. et al. (2002), "The open archival information system (OAIS) reference model and its usage", available at: <http://arc.aiaa.org/doi/pdf/10.2514/6.2002-T5-39> (accessed 9 May 2016).

Seles, A. (2016), "OAIS: its limitations and (how) can they be fixed", available at: <https://vimeo.com/165416927> (accessed 17 August 2016).