

Standards and Best Practices

To give you a bit of background, the basis of the way digital preservation is now practiced was developed in the 1990s and early 2000s. At that time, preservationists acknowledged that the number and variety of digital objects being created would overwhelm existing methods for managing them. To tackle this problem, multiple studies were conducted and initiatives started to address the lack of knowledge and methods to deal with these digital materials.¹ The best-known study is the 1996 *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*.² This report was generated by a task force created by the Commission on Preservation and Access (CPA) and the Research Libraries Group (RLG). The task force was charged to investigate contemporary roadblocks preventing the preservation of digital objects and to make recommendations on how to overcome these problems. One of the essential practices used in digital preservation programs today, engaging digital content creators as early as possible in the life of a digital object, was a result of this report. This practice includes educating creators on the long-term needs of digital objects—not just the technical needs, but also the need for contextual information to remain with the digital materials. This contextual information allows future users to interpret the original intentions of the creator and provides provenance that helps to boost the trustworthiness and authenticity of the objects.

There are a few ways for you to integrate this education into your organization's culture. The first is through a records management approach, where you require your content creators to use a limited set of software products for their tasks, mandating what format files will be saved in, requiring a specific folder and file structure with strict naming conventions, and so forth. This type of approach requires you to constantly communicate with, and in some cases supervise, your content creators. Many organizations are not able to allocate the resources necessary for this kind of oversight, and it usually works only for

internally produced content. Another approach is to work with creators at the point of content transfer. You could go through a standardized checklist with your creator to gain the essential contextual pieces needed to provide provenance and descriptive information to future users. This approach also allows you to limit the types of files your organization will receive by requiring content creators to migrate the files into standard, open source file formats before the transfer can be completed. You can add another layer to this *upon transfer* approach by providing education sessions to creators in your organization or to potential donors in the community you are trying to cultivate. This instruction can include recommendations for file formats, file-naming conventions, and tips for organization so that the transfer process, when it eventually occurs, goes more smoothly.

Preserving Digital Information had another pivotal recommendation—that a certification program for digital repositories be created so a network of trusted digital archives could be established. This recommendation led to two foundational international standards that the digital preservation community still relies upon today: the Open Archival Information System (OAIS) model and the Trustworthy Repositories Audit and Certification (TRAC) checklist.³

The OAIS model is a foundational document that digital preservationists use to discuss the nuts and bolts of a digital preservation repository. OAIS, ISO 14721:2012, was developed by the Consultative Committee for Space Data Systems (CCSDS) because the space industry produces an enormous amount of data that it is required by law to preserve and provide access to.⁴ The industry initially had no formalized plan for this data. The CCSDS realized at the outset that this standard would eventually be used beyond space data systems and that, even within its own industry, there was a tremendous variation in systems and technology. This led the CCSDS to develop the standard to be applicable across many different disciplines with many different technology requirements,

using language that is intentionally vague as to how to implement the standard. This approach makes the document extremely difficult to understand. In brief, OAIS describes an archival repository as a system that encompasses end-user needs, administrative oversight, the process by which digital materials become fully preserved and usable collections, and the foundational concept of packaging contextual information (metadata) with the digital objects throughout the entire process.⁵

To help you and other digital preservationists accomplish the goal of creating and maintaining a successful archival repository, OAIS defines several mandatory responsibilities for every digital preservation program. These responsibilities include what many archivists would consider basic practices of appraisal, arrangement and description, collection development policies, and access requirements. The appraisal requirements, for instance, specify that you have a donation agreement that defines what content is being transferred to the repository and the intellectual rights associated with the content, with particular emphasis on how intellectual rights intersect with preservation responsibilities.⁶

The arrangement and description aspects of the mandatory responsibilities require that you provide enough contextual information for the users to be able to independently discover and access all the content of the archive. To make digital content usable, you will often need to change the format or structure of the digital object. How the original document was formatted and structured is an essential piece of contextual information that needs to be recorded, as is the description of any changes you make. These arrangement and description responsibilities can be the most resource-intensive piece of the OAIS requirements, personnel-wise. With regard to collection development, OAIS requires that your digital archives program define who your end users are and what your users need from your archival repository. This will drive which kinds of digital objects you collect and how you preserve them.⁷

Finally, the access requirements, like the arrangement and description responsibilities, are more technology-focused than in traditional archival repositories, but with a similar emphasis on provenance and authenticity. Your digital archives program must have transparent policies and procedures to guarantee the long-term preservation of and access to your digital objects. Further, digital objects should be easy for your users to find. They should be provided to your users in a reliable manner, where the digital object provided to the user is an exact copy of the original the digital object in your repository or, if that is not possible, a copy of the original digital object in an updated format (also known as a *migrated* or *transformed* digital object) for the user to access. If you

provide an updated copy of the digital object, you should have available to your users an easy-to-understand audit trail that clearly indicates when the digital object was transformed, why it was transformed, how it was transformed, and who did the work.⁸

Beyond these mandatory responsibilities, OAIS also defines a model for building a digital preservation repository.⁹ This model defines a set of functions for how digital content is packaged and moved through a digital repository from content creator to end user and how the digital content is preserved over the long term. These functions include ingest, archival storage, data management, access, preservation planning, and administration. The first function, ingest, is a series of processes that define how a repository receives a Submission Information Package (SIP) from the content creator, how it validates that the transfer from the creator is uncorrupted and complete, how the SIP is transformed into an Archival Information Package (AIP), and how the AIP is transferred into preservation storage. The archival storage function includes more than the technology that stores the digital objects. It ensures that the digital content is unaltered (authentic) and readable in the long term. The archival storage function also emphasizes how important it is to monitor your preservation storage and plan for disasters. The next function, data management, is focused on the creation of, discoverability of, and documentation of the descriptive, preservation, and administrative metadata associated with your digital objects in your preservation system. The preservation planning function requires that your digital preservation program constantly monitor the digital preservation landscape, prepare for and implement changes as needed to keep your digital repository functional, and comply with international standards and best practices. The access function focuses on how users find and retrieve digital objects from your digital archive. Finally, the administration function defines how the day-to-day management of your digital preservation program is done.¹⁰ All of these functions can be developed in stages and then woven together to form the whole. You do not have to plan your program to be a fully compliant OAIS repository from the start. Instead, you should decide which function you are able to build out first and plan for that, leaving yourself the ability to integrate each new function together as you build them.

I place so much emphasis in this chapter on learning the OAIS standard because it is the common language that digital preservation professionals use to discuss repository development and maintenance with each other and with the information technology professionals who build and implement these systems. OAIS will soon be up for review, and it has been suggested that the wider digital preservation community, beyond the Consultative Committee for Space Data Systems, be allowed to suggest updates to make the

standard easier to read and more directly applicable to how repositories are currently functioning.¹¹

Digital repository developers needed an actionable way to answer the question “Is our repository OAIS-compliant?” For this, another ISO standard was created: 16363, Audit and Certification of Trustworthy Digital Repositories.¹² The development of this standard started when a working group comprised of members from the Research Libraries Group and the Online Computer Library Center authored a report in 2002, *Trusted Digital Repositories: Attributes and Responsibilities*, which defined a trusted digital repository and recommended that there be a continued push for digital archives certification programs.¹³ The report provided other high-level recommendations about where more research was needed to refine digital preservation implementation strategies. The Research Library Group first partnered with the National Archives and Records Administration in 2003, and later with the Center for Research Libraries in 2005, to operationalize the recommendations from *Trusted Digital Repositories: Attributes and Responsibilities*. These efforts resulted in the Trustworthy Repositories Audit and Certification (TRAC) checklist, published in 2007.¹⁴ This checklist was used as the basis for ISO Standard 16363 which is one of the certification methods used to determine a Trustworthy Digital Repository.¹⁵

While ISO 16363 is the formal standard, many digital preservation programs use the original 2007 TRAC report as a planning, self-assessment, and external evaluation tool instead of going through the formal certification process.¹⁶ TRAC was created through an international effort with contributors from different types of organizations that have a stake in the standards by which digital preservation programs are judged as consistent with recommended practice. These organizations included many entities beyond those that would traditionally be considered archival institutions, such as data repositories and research communities. This is an acknowledgement of the fact that digital preservation is most successful when content creators are involved with the effort as early as possible.

The TRAC document is an essential assessment tool because it emphasizes all aspects of a digital preservation program: technical setup, administrative policies and procedures, financial sustainability, and more. These aspects are split into three categories: organizational infrastructure, digital object management, and infrastructure and security risk management. This tool can be intimidating to first-time users due to its length and jargon-heavy language. The document was written with an assumption that the audience consists of professionals already familiar with digital preservation practice. However, each requirement is broken down into small, bite-sized pieces with

suggestions for how the repository can demonstrate achievement. The document was intentionally developed to be flexible so that it could be used by many different types of institutions. The document emphasizes that the assessment of an institution should be based upon that institution’s “mission, priorities, and stated commitments.”¹⁷ A caveat to this is that “regardless of the size, scope, or nature of the digital preservation program, a trusted repository must demonstrate an explicit, tangible, and long-term commitment to compliance with prevailing standards, policies, and practices.”¹⁸

There is a simpler, easier-to-understand certification process called the CoreTrustSeal, which has been specifically developed for data repositories.¹⁹ New digital preservation programs can use the requirements for planning purposes, and existing repositories can use the certification as a self-assessment tool. While TRAC has over one hundred requirements, the CoreTrustSeal has sixteen. The language of the CoreTrustSeal is data-focused, but by replacing the word *data* with *content* or *digital objects*, it is easy to see how these same requirements can be used to evaluate a digital preservation program. This certification program requires documentation of policies, procedures, licenses, and plans be publicly available when possible in an effort to promote transparency in how data repositories are set up and run. This transparency is an essential part of how a repository is deemed trustworthy.

OAIS, TRAC, and CoreTrustSeal emphasize the importance of documentation for a digital preservation system. Part of this documentation is the metadata associated with digital content, often grouped into four categories: descriptive, administrative, technical, and structural. Descriptive metadata is information about the digital objects; administrative metadata is information about rights, provenance, and a preservation audit trail; technical metadata is information about how to access the digital objects; and structural metadata is information about how digital objects relate to each other when they belong to a set.²⁰ Practically, these categories often overlap—one piece of metadata may fit into one or all of these categories at once. OAIS specifically requires metadata in the form of Preservation Description Information (PDI), which should include provenance, reference, fixity, contextual, and access rights information, all which contributes to maintaining a digital object’s authenticity and therefore could be considered administrative metadata.²¹ In practical terms, there are two metadata standards that are essential to the preservation of and access to digital materials: Preservation Metadata: Implementation Strategies (PREMIS) and Metadata Encoding and Transmission Standard (METS), both maintained by the Library of Congress.

PREMIS was originally a working group formed

by the Online Computer Library Center (OCLC) and the Research Libraries Group in 2003 created to build upon the report *A Metadata Framework to Support the Preservation of Digital Objects*, written by the Preservation Metadata Framework working group in 2002.²² The report proposed thirty metadata elements that the PREMIS working group used to create a data dictionary and a set of XML schemas for implementing the dictionary in digital preservation systems. The PREMIS Data Dictionary focuses on developing and maintaining preservation metadata as a means of keeping digital objects viable, usable, understandable, and authentic.²³ The working group that developed PREMIS required most of the core metadata to be generated and processed automatically by the repository system. Like OAIS, the PREMIS Data Dictionary is meant to be implementation-agnostic. Therefore, the way each digital preservation program produces and analyzes PREMIS metadata can be unique. A repository can comply with PREMIS without using the XML schemas provided by the PREMIS working group to create the information. As long as a repository can export its preservation metadata and crosswalk it to the Data Dictionary, that repository is considered PREMIS-compliant. Most importantly, the PREMIS Data Dictionary was developed to be OAIS-compliant so that all metadata generated to comply with the PREMIS standard will also comply with OAIS PDI requirements.²⁴

METS was originally developed for cataloging digital library objects. Its purpose is to extend descriptive metadata to include structural metadata that describes the organization of the component parts of an object. METS also allows descriptive metadata to be enriched with technical metadata describing the software and hardware information relevant to the digital object and, when necessary, the digitization specifications for a digital object. The Digital Library Federation provided an XML document format for encoding METS information. This XML document format allows repositories to point to descriptive metadata and administrative metadata listed in an externally maintained system like an EAD finding aid or a MARC record so that these efforts do not have to be duplicated, saving valuable time and resources. One of the unique aspects of the METS document is the hierarchical map that links elements of the structure to content files and their associated metadata. The METS document also includes a behavior section that can associate executable actions with the content. While METS was originally created for digitized images in an online library platform, it has been modified and extended over the years to meet the needs of digital preservation programs.²⁵ Like PREMIS, there are tools available that can automatically generate METS metadata and package that metadata with the digital content to form OAIS information packages.

These standards, together with others not mentioned here, create digital preservation best practice. In fact, since the early 2000s, when these standards were initially created, few new standards have been developed. Instead, the digital preservation community has focused on the practical implementations of these abstract reference models. These collaborative efforts have led to multiple case studies and templates being made available to the existing and new members of the digital preservation community to help develop new programs and boost existing programs to the next level. Institutions that have resources to devote to the actualization effort, working in concert, have developed tools and repository systems for their own use and then made these available to the community as a whole to benefit smaller organizations. These standards can be intimidating, but implementing best practice to conform to the standards is possible. I will discuss how in the following chapters of this report.

Notes

1. Erin Baucom, "A Brief History of Digital Preservation," in *Digital Preservation in Libraries: Preparing for a Sustainable Future*, ed. Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019), 3–19.
2. Donald Walters and John Garrett, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information* (Washington, DC: Commission on Preservation and Access, 1996), ERIC, <https://eric.ed.gov/?id=ED395602>.
3. Baucom, "A Brief History of Digital Preservation," 5–6.
4. International Organization for Standardization, *Space Data and Information Transfer Systems – Open Archival Information System (OAIS) – Reference Model*, ISO 14721:2012 (Geneva, Switzerland: ISO, approved March 2003; reaffirmed September 2012).
5. Brian Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*, 2nd ed., DPC Technology Watch Series (Glasgow, Scotland: Digital Preservation Coalition, October 1, 2014), <https://doi.org/10.7207/twr14-02>.
6. International Federation of Film Archives, "Digital Preservation Principles," accessed June 5, 2019, https://www.fiafnet.org/images/tinyUpload/E-Resources/Commission-And-PIP-Resources/TC_resources/Digital%20Preservation%20Principles%20v2%200.pdf.
7. International Federation of Film Archives, "Digital Preservation Principles."
8. International Federation of Film Archives, "Digital Preservation Principles."
9. For a visual representation of information packages moving through the functional entities of the OAIS reference model, see National Archives of Australia, figure 2 in "Digital Preservation Policy," February 20, 2018, www.naa.gov.au/about-us/organisation/accountability/operations-and-preservation/digital-preservation-policy.aspx.

10. International Federation of Film Archives, “Digital Preservation Principles.”
11. Thomas C. Wilson, “Rethinking Digital Preservation: Definitions, Models, and Requirements,” *Digital Library Perspectives* 33, no. 2 (March 10, 2017): 128–36, <https://doi.org/10.1108/DLP-08-2016-0029>.
12. International Organization for Standardization, *Space Data and Information Transfer Systems – Audit and Certification of Trustworthy Digital Repositories*, ISO 16363:2012 (Geneva, Switzerland: ISO, approved February 2012).
13. Research Libraries Group and Online Computer Library Center, *Trusted Digital Repositories: Attributes and Responsibilities* (Mountain View, CA: RLG, May 2002).
14. “Trustworthy Repositories Audit and Certification: Criteria and Checklist, version 1.0,” RLG—National Archives and Records Administration Digital Repository Certification Task Force (Chicago: Center for Research Libraries and Dublin, OH: OCLC, February 2007), https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf.
15. Baucom, “A Brief History of Digital Preservation,” 7–8.
16. *Trustworthy Repositories Audit and Certification: Criteria and Checklist*, version 1.0, RLG—National Archives and Records Administration Digital Repository Certification Task Force (Chicago: Center for Research Libraries and Dublin, OH: OCLC, February 2007), https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf.
17. *Trustworthy Repositories Audit and Certification*, 9.
18. *Trustworthy Repositories Audit and Certification*, 10.
19. CoreTrustSeal, “Core Trustworthy Data Repositories Requirements,” v01.00, November 2016, https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf.
20. Edward M. Corrado and Heather Lea Moulaison, *Digital Preservation for Libraries, Archives, and Museums* (Lanham, MD: Rowman and Littlefield, 2014), 113-115.
21. Corrado and Moulaison, *Digital Preservation for Libraries, Archives, and Museums*, 127-131.
22. OCLC/RLG Working Group on Preservation Metadata, *A Metadata Framework to Support the Preservation of Digital Objects* (Dublin, OH: OCLC, June 2002), https://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf.
23. “PREMIS Data Dictionary for Preservation Metadata, Version 3.0,” PREMIS Preservation Metadata Maintenance Activity website, Library of Congress, December 14, 2018, <http://www.loc.gov/standards/premis/v3/index.html>.
24. Priscilla Caplan, *Understanding PREMIS*, revised by PREMIS Editorial Committee (Washington, DC: Library of Congress, 2009, rev. 2017), www.loc.gov/standards/premis/understanding-premis-rev2017.pdf.
25. “METS: An Overview and Tutorial,” Metadata Encoding and Transmission Standard (METS) website, Library of Congress, March 30, 2017, www.loc.gov/standards/mets/METSOverview.v2.html.