ISQ

INFORMATION STANDARDS QUARTERLY

SPECIAL ISSUE: DIGITAL PRESERVATION

DIGITAL PRESERVATION
METADATA STANDARDS

TRUSTWORTHY
DIGITAL REPOSITORIES

UNIFIED DIGITAL
FORMATS REGISTRY

AUDIO-VISUAL
DIGITIZATION GUIDELINES

DIGITAL PRESERVATION
EDUCATION

NISO
How the information world
CONNECTS

ROBIN L. DALE AND EMILY B. GORE

Process Models and the Development of
TRUSTWORTHY DIGITAL
REPOSITORIES

In the past 15 years, there has been much effort to address the long-term preservation of digital assets, including the establishment of standards, related guidance, and best practices. In this article, the authors will give an overview of process models for preservation, including OAIS, InterPARES, and the DCC Curation Lifecycle Model, and the relationship of those process models to the development of standards related to trustworthy repositories. A discussion of work towards developing standards and best practices to establish trustworthy repositories begins with the seminal documents *Preserving Digital Information* and *Trusted Digital Repositories* (TDR) continues through currently used de facto standards TRAC, DRAMBORA, and nestor, and concludes with certification-related standards emerging from the OAIS family of standards. Process models and their intersections with efforts to provide guidance and set standards for trustworthy repositories guide the work of practitioners charged with long-term digital asset management across many disciplines.

Process Models

Process modeling is the activity of representing processes of a community, often so that current processes may be understood, analyzed, and improved. Process models are typically descriptive, prescriptive, and explanatory. The development of process models often begins by looking at the way processes have historically been performed and improvements for efficiency and effectiveness were determined. Process models then establish rules and guidelines that lead to desired process performance and provide explanations about the rationale of processes.

Early discourse about digital preservation tended to focus on specific technological strategies for digital files, but left important issues unaddressed. In developing process models for digital preservation, the community was forced to model and document the entire context in which those digital files existed, revealing overarching requirements for the infrastructure, supporting information models, processes, and systems in which they exist.

Open Archival Information System (OAIS)

In the early 1990s, the Consultative Committee for Space Data Systems (CCSDS) initiated work aimed at developing formal standards for the long-term storage of digital data generated from space missions. As described by Lavoie, this work was initially hindered because in early research, the CCSDS found no widely-accepted framework that could serve as a foundation for standards-building activities: nothing that established shared concepts and terminology associated with digital preservation, characterized the basic functions constituting a digital archiving system, or defined the important attributes of the digital information objects towards which preservation efforts could be directed. In 1995, the CCSDS began development of a framework that would serve the broadest constituency possible, incorporating relevant work from communities outside of the space data community including the seminal work, *Preserving Digital Information*, from the Task Force on Archiving of Digital Information.

Since the release of the CCSDS' draft OAIS reference model in 1999, archival repository systems worldwide have used OAIS as a benchmark and as the chief process model for the preservation of digital assets. The reference model

provides a common conceptual framework describing the environment, functional components, and information objects within a system responsible for the long-term preservation of digital materials. OAIS as a process model does not prescribe standards or technical architectures for archives or repositories; rather it gives a framework for further, more granular standards development and establishes an ontology for communication among repositories.

In 2003, the *Reference Model for an Open Archival Information System* was formalized and published as ISO 14721, paving the way for the development of future digital preservation standards work. The OAIS included a Roadmap for follow-on standards which led to the development of related process models. Follow-on or related standards development emerged including the *Producer-Archive Interface Methodology Abstract Standard* (PAIS) and the *PREMIS Data Dictionary for Preservation Metadata*. An additional standard was planned for the "accreditation of archives" but because of ongoing, parallel work, it was agreed that RLG and National Archives and Records Administration (NARA) would take this particular topic forward.

InterPARES

While OAIS was being developed, a process model for the long-term preservation of electronic records, InterPARES, was also in development. InterPARES, the International Research on Permanent Authentic Records in Electronic Systems, focuses on a model for ensuring the preservation, accuracy, reliability, and authenticity of electronic records. In Phase 1 (1999-2001), InterPARES work included the development of activity models for the selection and preservation functions,

and created a framework for requirements for assessing and maintaining authenticity of electronic records. Benchmark requirements supporting the presumption of authenticity as well as baseline requirements supporting the production of authentic copies of electronic records were also developed during this phase and were documented in the InterPARES *Preserve Electronic Records* model. While ensuring compliance with the OAIS model, the *Preserve Electronic Records* model defines processes specifically related to the preservation and delivery of authentic electronic records, and focuses only on essential preservation-related tasks. In Phase 2 (2002-2007), InterPARES shifted focus to newer kinds of electronic records: those which are dynamic, interactive, and experiential. The goal was to develop understanding surrounding their creation, maintenance, and preservation. Additional developments in this phase included methods for creating, maintaining, and preserving accurate, authentic, and reliable records in the arts, sciences, and government. Phase 3 (2007-2012) is currently underway and focuses on the movement of theory into practice through constituent adoption and education.

Digital Curation Centre Curation Lifecycle Model

A more recent model, the DCC (Digital Curation Centre) Curation Lifecycle Model, provides a graphical overview for the successful curation and preservation of digital assets from concept or receipt. The model aims to illustrate the steps or high-level processes necessary for long-term preservation, and is designed to be used in conjunction with relevant standards to plan curation and preservation activities to different levels of granularity. The DCC asserts that the

A **"trusted digital repository"** is one whose mission is to provide reliable, long-term access to managed digital resources for its designated community, now and in the future.

This concept is based on two major requirements:

1. the repository with associated policies, standards, and technology infrastructure will provide the framework for doing digital preservation, and

2. the repository is a trusted system, i.e., a system of software and hardware that can be relied upon to follow certain rules.

lifecycle model is intended to complement other models, like OAIS and InterPARES. Because of its intentional high-level overview, "workflow design, management issues, identification of processes and use of best practice can all be enhanced through the application of standards such as OAIS." The model defines three levels of preservation actions: full lifecycle, sequential, and occasional and points to the adherence of established best practices and standards for all levels of action. The DCC encourages use of the model as a training tool for data creators, data curators, and data users; to organize and plan resources; and to help organizations identify risks to their digital assets and plan management strategies for their successful curation.

From Process Models to Certified Digital Repositories

Having set the stage for the development of digital preservation frameworks and process models, *Preserving Digital Information* arguably also sets the stage for "trustworthy repositories" in its seminal work. "A critical component of the digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating, and providing access to digital collections…" In its recommendations, the Task Force articulated a need for "a process for certification of digital archives…to create an overall climate of trust about the prospects of preserving digital information." At the time, two potential models were recognized: an audit model based on those used to certify official depositories of government documents and a standards model where "participants claim to adhere to standards that an appropriate agency has certified as valid and appropriate; consumers then certify by their use whether the products and services actually adhere to the standards." Yet formal standards and well-accepted practices for digital preservation were slow to develop in the five years following the publication of *Preserving Digital Information*. Those that did emerge tended to be opposite ends of the standards spectrum: high-level process models and frameworks (OAIS, InterPARES) or more granular standards that addressed core parts of the digital preservation process (PAIS, PREMIS, etc.). The process models lacked the granularity required for an auditable certification process; individual, emerging standards lacked a framework for what constituted a trustworthy repository; and the community remained unable to come to a collective agreement on an exact definition of "trusted archives" as called for by the task force.

Defining Trustworthy Digital Repositories (TDRs)

In March 2000, RLG and OCLC began work to establish attributes of a digital repository for research organizations, building on and incorporating the then-emerging OAIS reference model. Representatives from libraries, archives, and data archives were charged to reach consensus on the characteristics and responsibilities of trusted digital repositories for large-scale, heterogeneous collections held by cultural organizations. The resulting work, *Trusted Digital Repositories: Attributes and Responsibilities*, articulated a framework of attributes and responsibilities for trusted, reliable, sustainable digital repositories capable of handling the range of materials held by large and small research institutions. It also defined a "trusted digital repository" as one whose mission is to provide reliable, long-term access to managed digital resources for its designated community, now and in the future. Inherent in this definition is the concept that preservation and access are inextricably linked but the framework was broad enough to accommodate different situations, architectures, and institutional responsibilities.

Jantz and Giarlo noted that a particular value of the TDR report was the concept that a "trusted digital repository" was based on two major requirements: "1) the repository with associated policies, standards, and technology infrastructure will provide the framework for doing digital preservation, and 2) the repository is a trusted system, i.e., a system of software and hardware that can be relied upon to follow certain rules." The

trusted system concept—that long-term digital preservation could not occur in a vacuum but instead existed within a larger organizational ecosystem that played key roles, as well as represented key vulnerabilities in the process—was an important step towards identifying trustworthy aspects of digital repositories. The document proved useful for institutions grappling with the long-term preservation of cultural heritage resources and was used in combination with the OAIS as a digital preservation planning tool. As a framework however, the TDR report concentrated on high-level organizational and technical attributes and only discussed potential models for digital repository certification. It refrained from being prescriptive about the specific nature of rapidly emerging digital repositories and archives and instead reiterated the call for certification of digital repositories, recommending the development of a certification program and the articulation of auditable criteria.

Developing Metrics for Certification

In 2003, RLG and the National Archives and Records Administration (NARA) created a joint task force to specifically address digital repository certification. The goal was to produce certification criteria and delineate a process for certification applicable to a range of digital repositories and archives. The membership of the RLG-NARA Task Force on Digital Repository Certification reflected that diversity, with practitioner-members from each of those organization types. All were chosen because of their experience in building and managing digital repositories. Continuity with earlier efforts was ensured by including members who had played active roles in the development of the OAIS standard and TDR report.

Beginning from a base of practitioner experience and leveraging concepts from existing documentation and standards for related types of certification (the ISO 9000 family of standards relating to organization and system management; ISO 17799 for data security and information management systems; the US Department of Defense Standard DoD 5015.2 (2002) for Records Management Applications, and many others), criteria were established and vetted using an iterative process. After two years, an audit tool comprising 88 metrics had been shaped and was released in draft as *An Audit Checklist for the Certification of Trusted Digital Repositories*.

A valuable public comment period brought important suggestions for improvement to the *Audit Checklist*, including the call for not only characteristics of a trusted digital repository, but also ways in which the presence of the attributes can be demonstrated and their qualities measured (see Ross and McHugh). By its publication, potential complexities of a formal audit and certification process were highlighted and questions were raised about applicability for existing "digital archives" of content. At a time when most

digital repositories were in the developmental stage, there was arguably an equal if not greater need for a planning/development tool for trusted repositories than a need for formalized audit and certification of digital repositories. How could a "best practice" audit tool be used to encourage and direct repository development without overwhelming institutions with nascent repositories? Was the *Audit Checklist* necessary and relevant for all digital repositories? Should regional needs or laws drive the development of several checklist versions? Could the checklist be easily used for self-assessment? And how would or could the *Audit Checklist* be applicable to repositories and digital content services that were established long before the *Audit Checklist* was developed?

The task force and other organizations considered those questions. The result was not only significant redevelopment of the *Audit Checklist*, but the development of two additional audit and criteria tools by two other organizations. The final phase of certification standards development saw an increase in interest, organizational sponsorship, and organizational participation. Two additional certification activities went into development in 2004–2005 and led to the release of complementary metrics for trustworthy repositories. Additionally, the three groups worked together to produce principles for minimum requirements for trustworthy digital preservation repositories.

1 The nestor Catalogue of Criteria for Trusted Digital Repositories

In December 2004, the German *nestor* project (Network of Expertise in Long-term STOrage of Digital Resources) set up the *nestor* Working Group on Trusted Digital Repository Certification to define a first catalog of criteria for trustworthiness and to prepare for the certification of digital repositories in accordance with nationally and internationally coordinated procedures. The aim of the project was to "establish a net of trustworthiness" in which long-term digital archives can function in various environments by formulating criteria that could be used for a broad spectrum of digital long-term repositories. Similar to the goals of the RLG-NARA task force, there was also a desire to provide information and self assessment assistance with the design, planning, and implementation of digital repositories.

Beginning with a small-scale survey on recent standards and usage within digital repositories within German institutions, the working group followed up with a public workshop in June 2005 and an expert round table in March 2006. Version 1.0 of the *Catalogue of Criteria for Trusted Digital Repositories* was released in June 2006. Comprising abstract criteria, enhanced with examples and explanations, the *Catalogue of Criteria* encompassed international standards but focused on applications in Germany. The central concepts

The aim of nestor was to "establish a net of trustworthiness" in which long-term digital archives can function in various environments by formulating criteria that could be used for a broad spectrum of digital long-term repositories.

driving the criteria include trustworthiness, as well as the concept that implementation of any certification process is a multi-step process for repositories and must be iterative. The application principles developed by *nestor*—Documentation, Transparency, Adequacy, and Measurability—were later adapted along with the Digital Curation Centre's needs for evidence in the RLG-NARA task force work. Today, the *nestor Catalogue of Criteria for Trusted Digital Repositories* continues to be in use in Germany in concert with training tools developed by the working group.

2   DRAMBORA: Digital Repository Audit Method Based on Risk Assessment

Developed jointly by the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE), the *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA) is intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognize their strengths. Borne out of a DCC repository assessment project, the initial basis for assessment was rooted in TRAC audit metrics (see #3) but was designed specifically with self-assessment in mind. DRAMBORA is a methodology for self-assessment, encouraging organizations to establish a comprehensive self-awareness of their objectives, activities, and assets before identifying, assessing, and managing the risks implicit within their organization. This method and the accompanying tool focus on organizations willing to perform a self-assessment to get an overview of the risks in their organization.

DRAMBORA focuses on risk management and asserts that the role of the curator or repository manager is to manage risks. Now available as an online, interactive toolkit, DRAMBORA defines six stages within the risk management process. Through the process of self-assessment, repository managers become aware of shortcomings and greatest risks. A systematic process guides the auditor to identify risks

to long-term preservation of repository content, and then scores each risk as a product between the likelihood of the risk occurring with the impact associated with that event. Mitigation of the risks can then be prioritized in descending order of the score so that risks can be effectively managed.

3   TRAC: Trusted Repositories Audit & Certification: Criteria & Checklist

During the final phase of metrics development, the RLG-NARA Task Force on Digital Repository Certification was fortunate to obtain valuable alliances with the then-new Digital Curation Centre, as well as colleagues in Germany directing the *nestor* project. A critical alliance with the Center for Research Libraries (CRL) also emerged. In 2005, the Center for Research Libraries was awarded a grant by The Andrew W. Mellon Foundation to develop the procedures and activities required to audit and certify digital archives. The CRL Certification of Digital Archives Project worked closely together with the RLG-NARA task force to redevelop the audit metrics and provided critical opportunities to develop and test the audit process itself. This practical testing, along with the DCC test audits that led to the development of DRAMBORA, contributed greatly to filling the gaps identified in the earlier draft, *Audit Checklist for the Certification of Trusted Digital Repositories*.

The final version of *TRAC* was published in February 2007 with 84 criteria broken out into three main sections: Organizational infrastructure; Digital object management; and Technologies, technical infrastructure, and security. It provides tools for the audit, assessment, and potential certification of digital repositories; establishes the documentation requirements for audit; delineates a process for certification; and establishes appropriate methodologies for determining the soundness and sustainability of digital repositories.

It currently serves as a de facto standard for repository audit and is being actively used by organizations as both a planning and self assessment tool. Additionally, it continues to serve as the basis of further CRL audit and certification work, including the National Science Foundation-funded project, Long-Lived Digital Collections. Currently, two repositories of interest, Portico and HathiTrust, have agreed to undergo CRL audits. Based on its recent audit findings, CRL has certified Portico as a trustworthy digital repository for the CRL community.

Ongoing Standards Development for Trustworthy Digital Repositories

After the publication of TRAC, the CCSDS working group responsible for OAIS-related standards (now called Mission Operations and Information Management Services or MOIMS) shepherded the TRAC certification metrics back into the CCSDS/ISO standards process. The MOIMS Repository

Audit and Certification (MOIMS-RAC) Working Group has endeavored over the last three years to formalize repository audit and certification metrics and continue the growth of the OAIS family of standards as envisioned at the outset of the OAIS work. Currently, two major contributions are in the standards process:

➡ *Audit and Certification of Trustworthy Digital Repositories* (currently designated CCSDS 652.0-R1, October 2009) is a draft standard that articulates the audit and certification criteria for trustworthy digital repositories. It is in the balloting and revision process and expected to be released very soon as the new international standard for certification.

➡ *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories* (Draft Recommended Practice CCSDS 000.0-R-0, Red Book) is meant primarily for those setting up and managing the organization performing the auditing and certification of digital repositories. Currently, ISO/IEC 17021, *Conformity Assessment Requirements for Bodies Providing Audit and Certification of Management Systems*, is the international standard that prescribes criteria for audit and certification agencies' work. The new CCSDS standard will incorporate new requirements and guidance for agencies to be accredited as complying with ISO/IEC 17021 with the objective of auditing and certifying candidate Trusted Digital Repositories (TDR).

With the formalization of these two documents, the standardization process for trustworthy digital repositories will have completed its first cycle. Not unlike the DCC's Curation Lifecycle Model, this cycle of understanding and standardization will continue as an iterative process. With a stable base of a process model, relevant standards and best practices for individual parts of the process will continue to be developed as the community's experience with and expertise in digital preservation grows. | FE | doi: 10.3789/isqv22n2.2010.02

ROBIN L. DALE <robin.dale@lyrasis.org> is the Director of Digital and Preservation Services for Lyrasis and as former RLG Program Officer was the principal author of *Trusted Digital Repositories: Attributes and Responsibilities, An Audit Checklist for the Certification of Trusted Digital Repositories*, and TRAC. Dale also served as a member of the OAIS, PREMIS, and ANSI/NISO Z39.87, *Technical Metadata for Digital Still Images*, working groups.

EMILY B. GORE <egore@clemson.edu> is the Head of Digital Initiatives & Information Technology for the Clemson University Libraries. Gore is the current co-chair of the ALA ALCTS-PARS Digital Preservation Interest Group, a sustaining member of the MetaArchive Cooperative, and an active repository manager.

RELEVANT LINKS

An Audit Checklist for the Certification of Digital Repositories. RLG and NARA, August 2005.
www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2416.pdf

Catalogue of Criteria for Trusted Digital Repositories. nestor Working Group on Trusted Repositories Certification, December 2006.
files.d-nb.de/nestor/materialien/nestor_mat_08-eng.pdf

The DCC Curation Lifecycle Model
www.dcc.ac.uk/sites/default/files/documents/publications/DCCLifecycle.pdf

DRAMBORA
www.repositoryaudit.eu/

InterPARES Project
www.interpares.org/

Jantz, Ronald and Michael J. Giarlo. Digital Preservation Architecture and Technology for Trusted Digital Repositories. D-Lib Magazine, 11(6), June 2005.
www.dlib.org/dlib/june05/jantz/06jantz.html

Lavoie, Brian. The Open Archival Information System Reference Model: Introductory Guide. DPC Technology Watch Report 04-01, 2004.
www.dpconline.org/vendor-reports/download-document/91-introduction-to-oais.html

PREMIS Data Dictionary for Preservation Metadata
www.loc.gov/standards/premis/

Producer-Archive Interface Methodology Abstract Standard
public.ccsds.org/publications/archive/651x0b1.pdf

Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data System, January 2002.
public.ccsds.org/publications/archive/650x0b1.pdf

Rolland, Colette and C. Thanos Pernici. A Comprehensive View of Process Engineering. In: Proceedings of the 10th International Conference CAiSE'98. B. Lecture Notes in Computer Science 1413, 1998.
ftp://sunsite.informatik.rwth-aachen.de/pub/CREWS/CREWS-98-18.pdf

Ross, Seamus and Andrew McHugh. The Role of Evidence in Establishing Trust in Repositories. D-Lib Magazine, 12(7/8), July/August 2006.
www.dlib.org/dlib/july06/ross/07ross.html

Ross, Seamus and Andrew McHugh. Preservation Pressure Points: Evaluating Diverse Evidence for Risk Management.
www.repositoryaudit.eu/images/PreservationPressurePoints.pdf

Task Force on Archiving of Digital Information. Preserving Digital Information. Council on Library and Information Resources, pub 63, May 1, 1996.
www.clir.org/pubs/reports/pub63watersgarrett.pdf

Trusted Digital Repositories: Attributes and Responsibilities. RLG-OCLC, 2002.
www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf

Trustworthy Repositories Audit & Certification: Criteria and Checklist. OCLC and CRL, version 1.0, February 2007.
www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf