

The Development of a Standard for Digital Repository Certification
Bruce Ambacher
U.S. National Archives and Records Administration

The mission to preserve digital information is emerging in a variety of digital repositories including archives, libraries, data centers, and commercial information technology centers. Regardless of their core program mission all have come to realize a shared mission to preserve digital information for future use.

The problem is how to ensure that the digital information that is preserved for the long-term is accessible and usable for as long as it is preserved. This part of the mission is complicated because we lack a clear understanding of the digital preservation issues or of all of the consequences of decisions taken as digital custodians. The only certainty is that the technology is constantly changing and today's digital preservation standards are inadequate to address future needs.

The growth in the number of repositories claiming to be capable of preserving digital information in the 1990s was met by equal concern and skepticism about the credibility of some claims and frustration with the lack of any valid measure to apply to the repositories. The various studies cited below highlighted the need for a digital certification standard.

Both of the major co-sponsors of the digital certification effort, the National Archives and Records Administration (NARA) and the Research Libraries Group (RLG), have long

traditions of concern for the integrity of the content of digital records. NARA was the first national archives to collect and preserve digital records. NARA acquired its first archival digital records, “1440 Hours Under the Sea,” from NASA in April 1970.¹ Today, it has the largest collection of archival digital holdings in the world and is planning for exponentially larger holdings with its multi-year, multi-Exabyte Electronic Records Archives (ERA).²

From the earliest days as a custodial archives nearly seventy years ago, NARA has been concerned with establishing and conveying the authenticity of the records in its custody. For textual materials this can be conveyed through an authentication program that is accepted in the courts.³ For digital materials it is more difficult to authenticate individual records. The focus has shifted to certifying the process and the generic records produced by the process. ERA’s emphases on scalability and on standards reinforce this position.

RLG’s interest in authentic digital records is nearly as old. It has been working to set standards and best practices in access and preservation for thirty years.⁴ In 1994 RLG and the Commission on Public Access established a Task Force to investigate digital information. Two years later the task force produced its report: “Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.”⁵ The report framed the key problem facing digital technology repositories facing information refreshment – maintaining and proving the authenticity of the refreshed records.

The report had a major influence on shaping certain of the thoughts and approaches of the Consultative Committee on Space Data Systems. P2, the archiving committee, had just begun developing the Reference Model for an Open Archival Information System (OAIS). P2 was quite receptive to the Report's development of Content, Context, Provenance, and Fixity.

During the development of OAIS, NARA hosted two CCSDS open workshops – Digital Archives Directions (DADs) in 1998 and the Archival Workshop on Ingest, Identification and Certification the following year. Both sought to identify and encourage activity in the major issues surrounding digital information. The evolution of CCSDS's P2's thinking is evident in the title of the second workshop – the major issues were bringing information into repositories, persistently identifying that information and assuring the quality of the repository itself.

RLG followed up the 1996 joint task force report with a RLG Preservation Working Group on Digital Archiving. A major outcome of that working group's recommendations was the establishment of a Digital Repository Certification Working Group co-hosted by RLG and NARA. Working group members were selected first for their individual skills and second for their institutional affiliation. That said, the group represents the major institutional interests – the national archives of the United States; the national libraries of the United States, France and the Netherlands; several major research universities; NASA; a major research laboratory; the Internet Archives; and RLG and OCLC. All have an interest in the issues of digital preservation and digital integrity.⁶

The project goal is simply and clearly stated: devise a methodology to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections, capable of winning our “trust” in their capacity and ability to preserve the digital information. The challenge is to produce certification requirements, delineate a process for certification, and identify a certifying body (or bodies) that can implement the process.

A number of specific steps, processes, and projects were outlined to accomplish the task of developing a certification plan:

- identify certifying body or bodies,
- identify a timetable for execution and adherence,
- identify the frequency or cycle of certification,
- define the conditions for revocation,
- create technical models - if possible, create economic models for sustainability of independent certifying program/body, and
- create implementation scenarios.

These specifics were joined by less specific, but equally valid criteria such as compliance, conscientiousness, the ability to sustain an audit, and adherence to standards and guidelines.⁷

In the first phase the task force worked toward these goals through a number of defined, achievable steps and processes including:

- review recent literature and project reports for appropriate checklists and criteria
- review and address the applicability of existing certification options in other disciplines to this endeavor
- identify a list of certifiable elements (attributes, processes, functions, activities)
- create a standard certification process or a framework that can be implemented across domains or types of digital repositories.

The task force recognized existing programs in other, related fields could provide the working model for this assessment program. The group carefully examined the programs already operating in the Society of American Archivists' – the "Guide to Managing an Institutional Archives," the Historical Manuscripts Commission's – "Standard for Record Repositories," and the American Association of Museums' – "Museum Assessment Program." The latter program offered the best model of traditional accreditation and self-assessment to achieve certification of museums.

Initially, the task force accomplished its work through independent effort, e-mail and monthly teleconferences; the pace of progress proved too slow. This was corrected by having weekly teleconferences for more than a year to ensure more continuity and greater progress. This period was highlighted by an invitational workshop at the National Archives in Washington, DC, March 21-23, 2005. This major phase ended on August 31, 2005, with a draft set of guidelines and a draft audit checklist to be applied against digital repositories.

The project entered a second phase in July. The Andrew Mellon Foundation has underwritten a grant to the Center for Research Libraries to:

- test and refine the metrics
- audit three repositories using the draft checklists
- determine the best set of processes and corresponding costs and the funding system for a certification program.

Progress during the grant phase will be quicker because the RLG co-chair will be the paid project director and a principal investigator during the grant period, allowing a concentration of time no one was able to allocate during the first phase when all efforts were in addition to other normal duties.⁸ The certification checklist and criteria will be revised based on the field testing and then exposed to a larger audience for further review and comment.⁹

Currently the checklist is influenced by the OAIS framework and measures services and metrics by OAIS criteria with appropriate sets of questions and categories of responses for each domain. Within Organization the checklist examines Governance and Organizational Viability, organizational structure and staffing, procedural accountability and policy, financial stability, and contracts, licenses and liabilities.¹⁰

Part B focuses on Archival Functions, Processes and Procedures. This part of the checklist includes an extensive section on ingest and data maintenance activities focusing on ingesting Submission Information Packages and transforming them into Archival Information Packages. The checklist also addresses Archival Storage, Preservation

Planning and Data Migration. These checklist questions are designed to elicit the repository's support for this aspect of certification. Typical questions include:

B2.7 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.

B3.1 Repository has documented preservation strategies.

B3.2 Repository implements/responds to strategies for AIP storage and migration

B 3.7 Repository actively monitors AIP integrity.

The checklist also addresses the repository's support for open access to its data through its use of metadata and its access management activities including its Dissemination Information Package procedures.¹¹

Part C of the checklist focuses on the relationship of the repository to its Designated Community. The repository must have in place (or at least access to) procedures for monitoring changes in any Designated Community. This implies

- A mechanism for specifying a Designated Community – presumably defined by its Knowledge Base. This may be separable into one or more identifiable components.
- Procedures for monitoring changes – OAIS suggests surveys, formal reviews, workshops, and individual interactions.
- The above should be done, where possible in collaboration with the producers and consumers.

- Changes in any “component” community should flag changes wherever appropriate.

The repository must itself have, or have access to, the ability to track digital technology and associated standards. This implies

- A procedure for tracking changes affecting functions or services of the repository.
- Participation in the tracking and in registry, or not.
- Involvement in prototype activities, or not.¹²

Preservation is a crucial activity for a certified digital repository whose core mission is first to ensure understandable information is preserved in an archival form suitable for long-term preservation, second, to provide long-term management, and third, to ensure that sufficient information is available to locate and retrieve the information over time.

An important component of the digital certification program is the OAIS concept of the Designated Community and the ability of the repository to meet the preservation and access requirements of the Designated Community or Communities it serves. To do so, the information it preserves must be independently understandable by the Designated Community without resort to the producer. This is achieved by the repository having appropriate technology, metadata, access services, and monitoring mechanisms. Where the repository does not acquire sufficient resources it must create them in the form of supplementary hardware, software, metadata, documentation and other forms of expertise. The success of this metric is dependent upon explicit written agreements with

the producer, and documented access policies readily know to the Designated Community.¹³

The final major section of the audit checklist addresses Technologies and Technological Infrastructure. It examines Archival Storage and the Infrastructure and processes associated with it. The checklist is seeking to ensure the preservation of Archival Information Packages for the long term and addresses system infrastructure, system security and disaster planning.¹⁴

The Task Force still has to resolve several significant issues including whether “best practices” associated with certification vary by domain, exactly which attributes will be certified, whether an official certifying body is needed and if so who will designate such a body, how people will be qualified to be certifiers, and the exact role repositories will play in developing or revising the certification criteria. Equally important is developing a certification process that both has broad based repository support and establishes a standard that is attainable by all who rightly deserve identification as a “certified repository.”

The draft *Audit Checklist for the Certification of Trusted Digital Repositories* is available for public review between September 1, 2005 and January 15, 2006. During this same timeframe the Checklist is being field tested at one digital repository. In January the drafting committee will resume its work and review and incorporate as appropriate both the public comments and the results of the field test in the final version of the Checklist.

NARA and RLG intend for the results to go into the standardization process through the International Organization of Standardization (ISO) Archiving series.

1. Accession Number NN-371-13, Record Group 255, in the records of the Electronic and Special Media Records Division, NARA; Bruce Ambacher, ed., *Thirty Years of Electronic Records*, Lanham, MD (Scarecrow Press: 2003), p. 1, 45.
2. www.archives.gov/electronic_records_archives/index.html.
3. Herman Kahn, "A Note on the Authentication of Documents," *American Archivist* XII (1949) p. 361-365.
4. www.rlg.org.
5. Donald Waters and John Garrett, *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*, 1996.
6. The complete membership list is available at: www.rlg.org.
7. www.rlg.org.
8. www.rlg.org; www.crl.edu., Press release: "Funding for Certification of Digital Archives Project Awarded to the Center for Research Libraries," May 3, 2005.

9. The repositories and collections to be audited are the Portico archives of E-journals maintained by Ithaka Harbors Inc., the archives of Elsevier journals maintained by the Koninklijke Bibliotheek, and a major archives of medical journals.

10. NARA-RLG Digital Repository Certification Task Force draft *Audit Checklist for the Certification of Trusted Digital Repositories*, Section A.

11. NARA-RLG Digital Repository Certification Task Force draft *Audit Checklist for the Certification of Trusted Digital Repositories*, Sections B1 –B3.

12. NARA-RLG Digital Repository Certification Task Force draft *Audit Checklist for the Certification of Trusted Digital Repositories*, Sections B4-B6.

13. NARA-RLG Digital Repository Certification Task Force draft *Audit Checklist for the Certification of Trusted Digital Repositories*, Section C.

14. NARA-RLG Digital Repository Certification Task Force draft *Audit Checklist for the Certification of Trusted Digital Repositories*, Section D.