

CERTIFICATION OF DIGITAL ARCHIVES – A BRIEF HISTORY AND STATUS REPORT

John Garrett ¹, David Giaretta ², Robert R. Downs ³, Steve Hughes ⁴, Mark Conrad ⁵, Bruce Ambacher ⁶, Simon Lambert ⁷, Helen Tibbo ⁸

(1) Garrett Software, 10703 Cordage Walk Columbia, MD 21044, USA,

(2) Giaretta Associates, UK

(3) CIESIN, Columbia University, USA

(4) NASA/JPL, Pasadena, CA, USA

(5) National Archives and Records Administration, Rocket Center, WV, USA

(6) NARA (retired), College Park, MD, USA, University of MD (retired), College Park, MD, USA

(7) STFC Science and Technology Facilities Council, UK

(8) University of North Carolina, Chapel Hill, NC, USA

ABSTRACT

The Reference Model for an Open Archival Information System (OAIS) standard is almost universally recognized and since it was originally published in 2002, it is a veritable grey beard of the digital archiving community. Most digital archives claim to be compliant with OAIS and in fact most are making capable and good intentioned efforts to do so. However, managers and funders of the archives may want those claims verified by outside observers.

There are a number of review processes that aim to offer some verification of the claims. However there is only one, the ISO process, which allows funders to have the confidence that a repository has been evaluated using the same well established procedures and degree of cross-checking that we all rely on in much of the rest of our lives, from the safety of our food to the security of our personal data held by governments.

Recently two additional archival standards were published. ISO 16363:2012 provides a set of specific metrics to verify adherence to OAIS intents. ISO 16919:2014 specifies how audits should be conducted and the competences that audit teams need to allow them to make consistent and reliable judgements about the capabilities of repositories. Together, these standards ensure that these audits are carried out using ISO's long standing and internationally recognized auditing process.

This paper provides a brief history of the two standards just mentioned and identifies a number of related standards. It also describes the efforts that are underway to develop nationally recognized digital repository auditing bodies and the individual auditors those bodies would require. Initial contacts have been made with a number of ISO National Bodies in order to work with them to accredit the auditors and the companies performing the audits. In the meantime, several of the authors of the above standards, through the Primary Trustworthy Digital Repository Authorisation Body (www.iso16363.org) are conducting high-level classes aimed at professionals in current auditing organizations and existing archives that are interested in performing audits or undergoing an audit to obtain certification.

INTRODUCTION

A great volume and variety of data is being produced. Funders, both public and private, are keen to ensure that the funds they put into this are not wasted and provide as much value as possible. One important way to do that is to ensure that the data can be re-used, immediately and into the future.

Use and re-use of data requires that adequate metadata is created and accompanies the data. Secondly it requires that the data and metadata are preserved.

The first step is being addressed by other efforts, e.g. those of the Consultative Committee for Space Data Systems (CCSDS) and the Research Data Alliance (RDA) (discussed in another paper), are trying to put in place standards and tools which will guide and help the production of the metadata up to and during the production of the data. Then comes the second step once the data have been produced, the information must be handed over to a repository, but significant questions must be answered. Which repository should be used? Is the repository doing digital preservation well enough? Can it be trusted?

Many repositories make claims about their preservation and archival expertise. However, managers and funders of the archives (and of the data producers) may want those claims verified by external impartial observers.

International Organization for Standardization (ISO) standards now exist that provide for certification of trustworthy digital repositories. The remainder of this paper will provide background regarding the development of the CCSDS/ISO standards leading up to these archive auditing standards. Following that, the paper will mention other certification efforts and discuss the steps that are needed to complete the needed certification infrastructure.

BACKGROUND

Organizational interest in the issues surrounding long-term preservation of digital information has existed since organizations began producing digital information. Discussion of such issues continued to grow and in the mid 1990's a number of organizations began concerted efforts to define and build the extensive infrastructure needed to ensure long-term preservation of digital information.

One example, the Research Libraries Group (RLG) and the Commission on Preservation and Access (CPA) (now merged into the Council on Library and Information Resources (CLIR)) created the Task Force on Archiving of Digital Information in December 1994. The task force report, *Preserving Digital Information*, was published in 1996. The following statements were among the conclusions of that Task Force.

- 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.
- A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information."

In that same time frame, the Consultative Committee on Space Data Systems (CCSDS) began a series of projects addressing digital preservation issues. These projects resulted in some of the fundamental digital data preservation standards and the current ISO standards for audit and certification of digital archives. A short discussion of the CCSDS efforts will provide some insight into the current archive certification situation.

HISTORY OF CCSDS/ISO OAIS STANDARDS SUITE

CCSDS was established in 1982 by the major space agencies of the time as an international forum for discussion of common problems and development of recommendation related to data handling in the space environment. It performed in this stand-alone manner for a number of years and its data handling recommendations were widely valued and used within the international space community.

In 1990 CCSDS entered into a cooperative agreement with the International Organization for Standardization (ISO) whereby CCSDS would act as a standards development agency and completed and approved CCSDS recommendations would be fed into ISO for review in order to become ISO standards. To effect this, Subcommittee 13 (Space data and information transfer systems) was set up as part of ISO Technical Committee 20 (Aircraft and space vehicles).

Over the years many CCSDS data handling recommendations were submitted to ISO and were also approved as ISO Standards. Then in the mid 1990's, ISO asked CCSDS to develop standards for the long-term storage of digital data. A CCSDS working group was formed at that time dedicated to development of such standards. Certainly, space agency personnel possessed a great deal of knowledge regarding the preservation of digital data since the space domain had generated some of the earliest examples of such data and the importance of the data (and inability to reacquire such data) had made it imperative that the information be preserved. Thus by the time this CCSDS project started, several of these agencies already had in excess of 40 years of experience maintaining large amounts of diverse digital data during times of great technical change.

Even with this experience, as discussions progressed in the CCSDS working group, the enormity of the task was recognized and decisions were made to approach the problem in an incremental fashion. It was also clear that similar information problems, were already being addressed for non-digital data by the archival, library and cultural heritage communities. It was also clear that those same outside institutions were themselves beginning to look to the future and look towards the need to handle digital information. Therefore, an excellent decision was made to expand the project and to bring in experts from fields beyond the space domain. This was done in the hope that the standards developed could serve not just the space community, but could also be used in the wider community and that these joint standards would prove beneficial for all the communities involved. At the time, this was a radical action from the viewpoint of the CCSDS leadership, but in fact it has been enormously successful.

Several large conclaves of digital data experts were held to begin to define and partition this work. However, before much progress was made, it soon became clear that although the experiences and needs of the communities were similar, the nascent approaches of the communities varied and the terminology each community used differed, limiting progress.

Therefore the working group leadership made another significant breakthrough with the determination that the first work item would be the creation of a digital preservation reference model that would establish common terms, concepts, and a basic information model for long-term preservation. Each community could continue to use its own vocabulary, but within the CCSDS working group, the reference model would allow the participants to translate their concepts into terminology used in the CCSDS Recommendations.

The result of this first effort was the *Reference Model for an Open Archival Information System (OAIS)*. A draft recommendation was circulated in 1999 and received wide review in the archival community resulting in many suggested updates and requests for additional clarification. Updates were made and the OAIS was first published as an approved Recommendation by CCSDS in 2002 and subsequently as ISO Standard 14721 in 2003 following a further ISO review. Both CCSDS and ISO require that their standards be reviewed on a periodic basis. Thus a revision of the standard was published in 2012. It is likely that a new CCSDS task will be set up within the next couple of years to review and, if needed, will revise the OAIS Reference Model recommendation once again. Due to the extremely wide adoption of this reference model, we expect to maintain backwards compatibility with the usages and terminology. The aim would be to provide clarification and possibly to expand the ability of the OAIS Reference Model to adapt to changes in the digital preservation environment.

An OAIS archive is an archive consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. The OAIS Reference Model defines a set of responsibilities that must be performed by an OAIS archive. OAIS archives should also be able to map the concepts and terms they use to the terms and concepts in the document's information model. The model provides a framework for the understanding and increased awareness of archival concepts needed for long-term digital information preservation and access, and for describing and comparing architectures and operations of existing and future archives. A roadmap of needed additional digital archival standards was also included.

Users of the OAIS Reference Model should note that it was never intended to be a design document. It is simply a reference model, but the actual design of software and systems may differ significantly as long as the concepts and functions specified in the standard are addressed by the actual preservation

systems. Another note is that the OAIS Reference Model does not and cannot address every preservation issue. In fact, the OAIS Reference Model identified a number of additional archival issues where additional standardization was desirable.

Thus following the completion the OAIS recommendation, several other digital preservation process standards and archive auditing standards have been produced. Although there have been many projects that could have been pursued, resources were identified from the willing participants to pursue two main prongs of standardization to add to the OAIS standards suite. These two areas are ingest of data into the archive and certification of archives.

Ingest Standards: PAIMAS and PAIS

One of the first standard projects following completion of OAIS Reference Model was development of a recommendation to identify and define the structure for the relationships and actions needed for an information producer to successfully and efficiently submit their information to an OAIS archive. This resulted in creation of the *Producer-Archive Interface Methodology Abstract Standard (PAIMAS)*. PAIMAS looked at requirements from the initial time of contact between the producer and the archive until the information objects are received and validated by the archive. Four phases were identified – Preliminary Definition, Formal Definition, Transfer, and Validation. Within each of the phases, sub-phases and needed individual actions were noted. PAIMAS was initially published as a CCSDS Recommendation in 2004 and then as an ISO Standard in 2006.

A recent CCSDS standards project is the development of the *Producer-Archive Interface Specification (PAIS)*, which provides a standard method for formally defining the digital objects that will be included in transfers from an information producer and provides for effective packaging of those digital objects into Submission Information Packages (SIPs) sent to an archive. This CCSDS Recommendation was published in 2014 and the associated ISO Standard was published in 2015. An associated tutorial will soon be available from CCSDS.

CCSDS Audit and Certification Standards

TRAC – predecessor to CCSDS RAC Standards

During its development, the OAIS Reference Model was widely distributed for review and comment and attempts were made to hear and respond to wide swaths of the preservation community. As various community segments participated and contributed to the OAIS Reference Model, they tended to adopt its concepts and in many cases its terminology. So even before the OAIS Reference Model was published as a final approved CCSDS Recommendation, it was widely recognized. With this wide recognition, it was inevitable that organizations would soon start to claim conformance with OAIS.

While the creators of the OAIS were gratified by the wide acceptance of OAIS, they were also aware of the questions raised. What did conformance to OAIS really mean? Did conformance to OAIS mean that an archive could be trusted to preserve digital information for the long-term? If OAIS conformance did mean that a repository could be trusted with the preservation of digital information, how could those outside of the repository know if the repository was truly OAIS conformant? The CCSDS Working Group responsible for the OAIS Reference Model was well aware that these types of questions would arise. They planned for the development of archives auditing standards and in fact the OAIS Reference Model standard calls out the need for many OAIS-related standards including certification standards.

In parallel to the CCSDS PAIMAS effort mentioned above, CCSDS was able to delegate initial development of archive certification efforts to another organization with the intention that they would create a draft document that would then be submitted back the CCSDS Working Group for further development and subsequent international standardization. The organization that undertook this initial drafting effort was a joint task force set up by the US National Archives and Records Administration (NARA) and the Research Libraries Group (RLG) (now merged with the Online Computer Library Center (OCLC)). Several of the OAIS authors participated in the NARA-RLG Task Force on Digital Repositories helping to advance the certification efforts.

Significant amounts of work were put into the development efforts. The task force interacted with other national and international organizations that were working on similar documents. The guidelines were developed iteratively and feedback was received from programs of test audits. This project reached its culmination when the joint task force published the Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) in 2007.

Note that TRAC is neither a CCSDS nor an ISO standard. At the end, TRAC was rather precipitously published as a final report of the task force before it disbanded. Following TRAC's publication, the TRAC document itself was no longer being updated and as planned the document was submitted to CCSDS and ISO for further development. In the interim until the standardization could be completed and a more complete certification infrastructure could be established, several organizations continued to make use of TRAC. The Center for Research Libraries (CRL) continues to assess organizations to TRAC and have so far CRL has identified six repositories that meet CRL's criteria for compliance with the TRAC report. Since TRAC was always intended as a temporary intermediate produce, now that ISO 16363 is completed, it seems likely that those organizations assessing against TRAC will upgrade their efforts to either provide consultation services for archives seeking ISO 16363 certification or will follow ISO 16919 procedures and become an accredited Conformance Body that is able to certify repositories as ISO 16363 Conformant.

CCSDS RAC Standards

In 2007, with the TRAC document as a starting point, CCSDS Repository Audit and Certification (RAC) Working Group began developing the *Audit and Certification of Trustworthy Digital Repositories* standard (also known as the TDR metrics document or ISO 16363). ISO 16363 provides a set of specific metrics to verify adherence to OAIS intents. ISO 16363 organizes 126 TDR specific assessment items into three areas of archival practice - organizational infrastructure, digital object management, and infrastructure and security risk management. Organizational infrastructure includes requirements for governance and organizational viability, organizational structure and staffing, procedural accountability and preservation policy framework, financial sustainability, and contracts, licenses, and liabilities. Assessments in the digital object management area include ingest: acquisition of content, ingest: creation of the AIP, preservation planning, AIP preservation, information management, and access management. Subcategories for infrastructure and security risk management are technical infrastructure risk management and security risk management.

ISO 16363 provided the metrics that could be used for certifying digital archives. Now the question became, how can we ensure that these metrics are used fairly and uniformly? Within ISO, general guidance for how audits are conducted is provided by ISO 17021 *Requirements for Audit and Certification of Management Systems*. ISO/IEC 17021-1:2015 contains general principles and requirements for the competence, consistency and impartiality of bodies providing audit and certification of all types of management systems, but it doesn't provide specific requirements regarding the application of these requirements to digital archives. A companion document that applies ISO 17021 to archives and makes use of ISO 16363 TDR metrics was also needed. That document is the *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories* standard (also known as the TDR auditing guidelines document or ISO 16919).

OTHER ASSESSMENT METHODS

DSA

The Data Seal of Approval (DSA) originated in the European social science and humanities communities but is intended to be applicable across all domains. The DSA is viewed by some as a preservation related assessment, but in fact is both more and less than that. Some of DSA's guidelines are related to information preservation, but, for example, other DSA guidelines are related to relationships and division of labor between the data repository and its producers and consumers. The DSA provides 16 quality guidelines. There are 3 guidelines that assess the practices of a data repository's producer, 3 that assess the practices of a data repository's consumers and 10 guidelines for the repository itself. One guideline requires compliance with the OAIS tasks and functions. This different focus means that some organizations that have been awarded a Data Seal will not be able to

be ISO 16363 certified as a trustworthy digital. Conversely it is possible that an ISO 16363 certified trustworthy archive would not meet some DSA guidelines that are not related to trustworthiness and long-term preservation.

The original version of the DSA guidelines was published in 2010 and revised version was published in 2013. Repositories seeking to obtain the DSA assess themselves against the 16 DSA guidelines and then pass on the documentation to the DSA staff. There is no audit process or site visit. However, the self-assessment materials are reviewed by an external peer. If approved by the peer reviewer, the self-assessment is published and the repository is allowed to display the DSA seal. The DSA seal is valid indefinitely, but the date the DSA was awarded is noted. If repositories want to comply with updated DSA guidelines, they need to update their assessments. About 50 organizations or activities within organizations have been awarded the Data Seal of Approval.

The DSA, with its low costs, may be of value to smaller archives. However, the lack of a rigorous process including an on-site outside auditor and mandated periodic recertification requirements, will not be adequate for the demands of many larger communities and their funders and for those archives the international ISO 16363 certification should be sought.

DIN

The Deutsches Institut Fur Normung E.V. (DIN) (the German Standards Body) working group on Trustworthy Digital Archives published DIN standard 31644 *Information and Documentation - Criteria for Trustworthy Digital Archives* in 2012. It was based on a set of 10 core requirements developed by an international collaboration of archives, libraries and museums to all institutions who aim to preserve information in digital form. The scope has been broadened to all institutions who aim to preserve information in digital form.

The main text of the standard consists of 34 requirements structured in three parts – organization, management of intellectual entities and their representations, and infrastructure and security. Basically the same structural divisions in ISO 16363 and TRAC. A useful feature of DIN 31644 is a set of appendices with examples of digital archives and best practices for each requirement. Few, if any, organizations have been third-party certified as complying with DIN 31644. If organizations are DIN 31644 certified with its national scope, it is likely a small step to also become ISO 16363 certified and thereby receive the international recognition inherent with certification to an ISO standard using ISO processes.

The European Framework of Audit and Certification

The European Framework of Audit and Certification identified three levels of increasing trustworthiness.

1. Basic Certification is granted to repositories which obtain DSA certification;
2. Extended Certification is granted to Basic Certification repositories which provide a publicly available and externally reviewed self-audit based on ISO 16363 or DIN 31644;
3. Formal Certification is granted to Basic Certification repositories which also obtain a full external certification based on ISO 16363 or a DIN 31644 equivalent.

It was believed that the various levels with the framework allowed both large and small organizations and mature archives and less established organizations an ability to demonstrate their level of trustworthiness to their communities. However, we are not aware of archives who have committed resources to obtaining multiple certifications. For example, it seems that an archive hoping to obtain third party certification to ISO 16363 will simply apply their resources to a process of self-audit, followed by outside review of that self-audit and finally undergoing a third party audit.

ISO CERTIFICATIONS AND ACCREDITATIONS

ISO is an independent, non-governmental, membership organization that publishes international standards. Each of the 150+ member countries of ISO is represented within ISO by a single National Body representing that country. Other standards developing organizations, such as CCSDS, may work

in liaison with ISO to produce ISO standards. ISO itself is responsible for the development of standards but it is not directly responsible for providing certifications or conformity assessments.

End users need to obtain their certification of compliance to ISO standards from external third party certification bodies. Typically certification bodies need to be accredited to perform conformity assessments and award certifications. Again ISO itself does not provide accreditations. Typically regional or national organizations are set up to accredit the certification bodies in each region or country. To ensure uniform accreditation and certification activities worldwide, an ISO sister organization, the International Accreditation Forum (IAF), has been set up to coordinate with national and regional organizations and to provide worldwide recognition of certifications granted.

These National Bodies participate in the development of ISO standards. Sometimes the national bodies also participate in accreditation and certification activities related to ISO standards. In other countries, other organizations, or sister organizations to the national bodies, are set up to handle the accreditation and certification activities.

STATUS OF NATIONAL CERTIFICATION EFFORTS

As the CCSDS/ISO RAC standardization efforts were drawing to a close, several participants of the CCSDS RAC Working Group formed the Primary Trustworthy Digital Repository Authorisation Body (PTAB) to provide training regarding usage of OAIS, ISO 16363 and ISO 16919 and to work towards development of a worldwide archival auditing infrastructure. Progress continues. PTAB is offering high-level ISO 16363 training courses. Among others, courses have been presented in London, the Hague, Washington, DC, Pasadena, CA, and one is planned in Rome. Additional courses will be scheduled as availability and demands dictate. For more information on PTAB or the course, visit <http://www.iso16363.org>.

Progress is also being made in developing programs and structures for repository certification. Italian legislation requires that any archive that maintains data from others needs to be certified. The criteria used are written into Italian law, but many are essentially the same requirements as those in ISO 16363. They have set up their own certification processes and infrastructure.

Efforts are underway in several other nations to set up the accreditation of organizations to perform audits to ISO 16363 using ISO procedures and infrastructures. Efforts are progressing along different paths and at different speeds within different countries. We expect that once the efforts in one country are complete, it will provide a template that will be followed in other countries. These efforts have probably progressed the furthest within the USA.

Within the US, the ISO member body is the American National Standards Institute (ANSI), which participates in and coordinates with other U.S. organizations that participate in the development of ISO standards. The main organization responsible for accreditation and certification activities is a sister organization set up jointly by ANSI and the American Society for Quality (ASQ). It is the ANSI/ASQ National Accreditation Board (ANAB). The ANAB portfolio includes accreditation of organizations performing audits for many types of management systems and products. The ANAB portfolio also includes ANSI programs for personnel certifiers, which may eventually include certification of auditors.

Within the US, ISO 16363 is being recognized and has been endorsed by many organizations, including the Society of American Archivists, the oldest and largest North American national archival professional association. As a result of the rising demand for certification of trustworthy archives, ANAB is developing an accreditation program for certification of organizations that conduct archive certification programs according to ISO 16919, to ensure that the archives conform to ISO 16363 and is beginning to offer related training sessions (www.anab.org).

The authors and PTAB members would welcome discussions related to development of accreditation and certification bodies within other countries and are willing to work together and provide training.

REFERENCES

ANAB (2015), *Prelaunch Announcement ISO 16363 Certification Program*. Available at: <http://anab.org/media/53684/hu319.pdf>

CCSDS (2012), *Reference Model for an Open Archival Information System (OAIS)*. Recommendation for Space Data System Practices, CCSDS 650.0-M-2. Washington, D.C.: June 2012. Issue 2. [Equivalent to ISO 14721:2012.] Available at: <http://www.ccsds.org/publications/archive/650x0m2.pdf>

CCSDS (2011), *Audit and Certification of Trustworthy Digital Repositories*. Recommendation for Space Data System Practices, CCSDS 652.0-M-1. Washington, D.C.: September 2011. Issue 1. [Equivalent to ISO 16363:2012] Available at: <http://www.ccsds.org/publications/archive/652x0m1.pdf>

CCSDS (2014), *Guidelines for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories*. Issue 1. Recommendation for Space Data System Practices, CCSDS 652.1-M-2. Washington, D.C.: March 2014. Issue 2. [Equivalent to ISO 16919:2014] Available at: <http://www.ccsds.org/publications/archive/652x1m2.pdf>

CCSDS (2004), *Producer-Archive Interface Methodology Abstract Standard (PAIMAS)*. Recommendation for Space Data System Practice, CCSDS 651.0-M-1. Washington, D.C.: May 2004. Issue 1. [Equivalent to ISO 20652:2006.] Available at: <http://www.ccsds.org/publications/archive/651x1m1.pdf>

CCSDS (2014), *Producer Archive Interface Specification (PAIS)*. Recommendation for Space Data System Standards, CCSDS 651.1-B-1. Washington, D.C.: February 2014. Issue 1. [Equivalent to ISO 20104:2015.] Available at: <http://www.ccsds.org/publications/archive/653x0b1.pdf>

Data Seal of Approval. Available at: <http://www.datasealofapproval.org/en/>

Deutsches Institut Fur Normung E.V. (2012), DIN 31644 *Information and documentation - Criteria for trustworthy digital archives*. Original language German Available at: <http://www.beuth.de/en/standard/din-31644/147058907?SearchID=966013709>

ISO/IEC (2011), *Conformity Assessment—Requirements for Bodies Providing Audit and Certification of Management Systems*. 2nd ed. International Standard, ISO/IEC 17021:2011. Geneva: ISO, 2011. Available at: http://www.iso.org/iso/home/store/catalogue_ics.htm

PTAB (2014), *PTAB Training Program*. Available at: <http://www.iso16363.org/courses/>

Task Force on Archiving of Digital Information (1996), *Preserving Digital Information*. Available at: <http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>

Task Force on Trusted Digital Repository (2007), *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*. Available at: http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf