Audit and Certification Process for Digital Repositories

David Giaretta [1], Mark Conrad [2], John Garrett [3], Terry Longstreth [4], Simon Lambert [5], Barbara Sierman [6], Steve Hughes[7], Helen Tibbo[8]

*[1] **STFC/APA***
Millers Cottage, Yetminster, Dorset, UK
EMail: david@giaretta.org
*[3] **ADNET Systems/NASA***
GSFC, Greenbelt, MD, USA
EMail: John.G.Garrett@nasa.gov
*[5] **STFC***
Rutherford Appleton Lab., Oxon OX11 0QX, UK
EMail: simon.lambert@stfc.ac.uk
*[7] **JPL***
Pasadena, California 91109, USA
Email: john.s.hughes@jpl.nasa.gov

*[2] **NARA***
Rocket Center, WV 26726, USA
EMail: mark.conrad@nara.gov
*[4] **Retired***
Laurel, MD, USA
EMail: longstreth@acm.org
*[6] **KB***
The Hague, The Netherlands
EMail: Barbara.Sierman@KB.nl
*[8] **Consultant***
Chapel Hill, NC
Email: tibbo@email.unc.edu

ABSTRACT

This paper provides details of the process by which digital repositories can be formally evaluated in terms of their ability to preserve the digitally encoded information with which they have been entrusted. The ISO standards on which this is based will be described and explained as will the structures at European, US and global levels which can provide the ISO audit service. The relationship of these standards to OAIS and the earlier TRAC document is also described.

A European framework for audit and certification of trusted repositories which includes three levels of certification is described. This three level process should provide an easy way for repositories to start the process and proceed to the highest level: the ISO audit. .

A number of test audits have been conducted with repositories in Europe and the USA. The lessons learned from these test audits and the way in which they have been incorporated into the audit and certification process is described.

At the time of writing the capacity in terms of number of repositories which can be audited per year is small. Plans for increasing that capacity will be described as will the qualifications which auditors will be required to have. We also describe the decisions which have been taken to minimise the effort (and hence the cost) required both from the repository and from the auditors, as well as the tools which may assist in this aim.

Finally an assessment of the possible impact of this type of audit and certification on the practice of preserving digital information will be given.

Keywords: OAIS, digital preservation, audit, certification

INTRODUCTION

The Preserving Digital Information report of the Task Force on Archiving of Digital Information [1] declared,

- a critical component of 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.

The issue of certification, and how to evaluate trust into the future, as opposed to a relatively temporary trust which may be more simply tested, has been a recurring request, repeated in many subsequent studies and workshops.

The challenge was then to create a mechanism to satisfy this demand. This paper provides an insight into how this is being done.

TESTABILITY AND KEY OAIS CONCEPTS

The Reference Model for an Open Archival Information System (OAIS) [2] is "now adopted as the 'de facto' standard for building digital archives" [3].

An important principle from the OAIS standard is the need for claims about preservation of digitally encoded information to be testable. These are summarised next – in what follows OAIS terms are in bold and capitalised.

Preservation

We need first some methodology by which to test the basic claim that someone is preserving some digitally encoded information; without such a test this is a meaningless claim. OAIS introduces the, quite reasonable, test that the digital object must somehow be useable and understandable in the future. However by itself this is too broad - are we to be forced to ensure that the digitally encoded designs of a battleship are to be understood by everyone, for example a 6 year old child? In order to make this a practical test the obvious next refinement is to describe the type of person - and more particularly their background knowledge - by whom the information should be understandable. Thus OAIS introduces the concept of **Designated Community**, defined as an identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. Note that a Designated Community is defined by the repository and this definition may change/evolve over time.

Bringing these ideas together we can then say, following OAIS, that preserving digitally encoded information means that we must ensure that the information to be preserved is **Independently Understandable** to (and usable by) the Designated Community.

We are clearly concerned about long term preservation, but how long is that? OAIS defines **Long Term** as long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing Designated Community. Long Term may extend indefinitely

Definition of the Designated Community

An important clarification is needed here, namely that the definition of the Designated Community is left to the repository. The same digital object held in different repositories could be being preserved for different Designated Communities, each of which could consist of many disjoint communities.

The quid pro quo is that those funding the repository, or entrusting their digital objects to the repository, can judge whether the definition of the Designated Community is appropriate for their needs.

OAIS Conformance

The OAIS standard itself defines conformance in terms of the Information Model and the mandatory responsibilities.

OAIS introduces a number of important concepts and conformance criteria; however this is not enough on which to base a certification scheme. The next section describes how, using OAIS as a basis, such schemes have been attempted.

TRAC and related documents

Section 1.5 of OAIS (the section entitled *Road map for development of related standards*) included an item for accreditation of archives, reflecting the long-standing demand for a standard against which Repositories of digital information may be audited and on which an international accreditation and certification process may be based. It was agreed that RLG and NARA take a lead on this follow-on standard.

A group was gathered by NARA and RLG (the latter subsequently incorporated into OCLC) to form the Task Force on Trusted Digital Repositories. This group produced the Trustworthy Repositories Audit and Certification : Criteria and Checklist [4]. The work combined concepts from OAIS and the Trusted Digital Repositories: Attributes and Responsibilities [5]. The latter allowed the group to supplement OAIS with considerations of financial stability and training of personnel.

The document has a number of metrics grouped into

- Organisational Infrastructure
- Digital Object Management and Technologies
- Technical Infrastructure
- Security.

Accompanying each of the metrics is extensive additional explanatory text and examples of the types of evidence which might be used as proof of fulfilling the metrics. The document has been used as the basis for internal and test audits in a number of repositories, however it is not part of a formal audit and certification process.

Other work in this area includes:

- the German preservation consortium, nestor, produced in 2006 a Catalogue of Criteria for Trusted Digital Repositories [6]
- Ross et al [9] produced comments on the TRAC document in 2006
- in early 2007 representatives from the Digital Curation Centre (DCC, http://www.dcc.ac.uk), DigitalPreservationEurope (DPE, http://www.digitalpreservationeurope.eu/), NESTOR (Germany) and the Centre for Research Libraries (North America) met and produced a list of 10 core criteria for digital preservation repositories, to guide further international efforts on auditing and certifying repositories [7]. A comparison of this list with the OAIS responsibilities was produced in 2008 [8].
- a cross-walk between the TRAC, nestor and Ross documents was produced in 2007 [10]
- the DCC and DPE projects produced the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit. This toolkit is intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognise their strengths.

All this work has been helpful in providing information and experience in assessing digital repositories, and some provide a local or project-backed certificate of quality. However none provide an ISO based accreditation and certification system of the kind which is available in other areas, such as the one concerning Information Security based on ISO 27001 series. Without this we cannot expect to have a mark of quality and trustability for digital repositories which is recognised world-wide. Efforts to produce such a system are described next.

DEVELOPMENT OF AN ISO ACCREDITATION AND CERTIFICATION METRICS

The development of OAIS was hosted by the Consultative Committee for Space Data Systems (CCSDS, http://www.ccsds.org) and approved by ISO as ISO 14721. OAIS contained a roadmap which listed a number of possible follow-on standards, some of which e.g. the Producer-archive interface -- Methodology abstract standard (ISO 20652:2008), have already become ISO standards, after development within CCSDS.

The need for a standard for certification of archives was included in that list and the RLG/NARA work, described above, which produced TRAC was the first step in that process. The next step was to bring the output of the RLG/NARA working group back into CCSDS. This has been done and the Digital Repository Audit and Certification (RAC) Working Group [11] has been created, the CCSDS details are available from http://cwe.ccsds.org/moims/default.aspx#_MOIMS-RAC, while the working documents are available from http://wiki.digitalrepositoryauditandcertification.org. Both may be read by anybody but, in order to avoid hackers, only authorised users may add to them. The openness of the development process is particularly important and the latter site contains the notes from the weekly virtual meetings as well as the live working version of the draft standards.

Besides developing the metrics, which started from the TRAC document, the working group also worked on the strategy for creating the accreditation and certification process. As a result of the review of existing systems which have accreditation and certification standard processes it became clear that there was a need for two documents

1. **Audit and Certification of Trustworthy Digital Repositories** [12]
2. **Requirements for bodies providing audit and certification of candidate trustworthy digital repositories** [13]

The first document lists the metrics against which a digital repository may be judged. It is anticipated that this list will be used for internal metrics or peer-review of repositories, as well as for the formal ISO audit process. In addition tools such as DRAMBORA could use these metrics as guidance for its risk assessments.

Understanding the ISO Trusted Digital Repository Metrics

It is clear that one cannot cover all possible situations in the metrics, nor can one prescribe exactly what each repository must do. This is the case with all types of audits. Instead one must leave a lot to the judgment of the auditors.

To understand the way in which the metrics in ***Audit and Certification of Trustworthy Digital Repositories*** (referred to below as the "metrics document") were written it is helpful to think about the document in the following way, building it up in the same way that the authors of that document built it up.

A very important thing to understand is that in judging a repository one could look at many types of issues. For example is the restaurant good, is the lighting adequate, is there wheelchair access, does the repository respond to requests within 3 minutes, is it easy to find what one is looking for and so on. However these are **not** the things against which the repository is to be judged here. Instead we are concerned about how well a repository preserves the digitally encoded information with which it has been entrusted.

With this in mind, one could say that since the audit and certification depends on the judgment of the auditors, the metrics document could have one metric, namely "*Make sure the repository does a good job in preserving its holdings*".

Of course this would not be adequate. We need to provide more guidance for the auditors. Therefore we start by saying "Well at least look at the organisation – make sure it cannot suddenly go out of business, and also make sure that they know how to preserve the digital objects." This one can say that there are two guidelines for auditors:

- Look at the organisation and its finances
- Look at the way it takes care of the digital stuff

In fact there is a third area, which one could argue is part of the second one, namely:

- Make sure that the digital holdings cannot be stolen or otherwise lost.

The reason this third bullet is added is that the repository could undergo a security audit separately (ISO 27000) so that it seemed sensible to provide a separate group which could essentially be replaced by ISO 27000 certification – but such additional certification is definitely not required.

Therefore we have three main headings:

- **Organisational Infrastructure**
- **Digital Object Management**
- **Infrastructure and Security Risk Management**

Continuing this process we can specify the topics where the auditor really needs to be sure to look. The metrics document has the following breakdown:

- **Organisational Infrastructure**
  - **GOVERNANCE & ORGANIZATIONAL VIABILITY**
  - **ORGANIZATIONAL STRUCTURE & STAFFING**
  - **PROCEDURAL ACCOUNTABILITY & PRESERVATION POLICY FRAMEWORK**
  - **FINANCIAL SUSTAINABILITY**
  - **CONTRACTS, LICENSES, & LIABILITIES**
- **Digital Object Management**
  - **INGEST: ACQUISITION OF CONTENT**
  - **INGEST: CREATION OF THE AIP**
  - **PRESERVATION PLANNING**
  - **AIP PRESERVATION**
  - **INFORMATION MANAGEMENT**
  - **ACCESS MANAGEMENT**
- **Infrastructure and Security Risk Management**
  - **TECHNICAL INFRASTRUCTURE RISK MANAGEMENT**
  - **SECURITY RISK MANAGEMENT**

This breakdown into topics is not unique and indeed several different breakdowns have been tried; this one seemed to fit best.

Looking in even more detail the guidance for the auditors is further split out into metrics. Some of these metrics are broken into sub-metrics indicating that the auditor needs to check these even more specific points; a few of these sub-metrics have some even more specific sub-sub-metrics specified. Even these sub-sub-metrics are not hugely specific – it is still a matter for the judgment of the auditor. Indeed the metrics themselves are a matter of judgment; in this case of course it is the judgment of the working group which produced the metrics.

Since much depends on the judgment, the questions arise – who are the auditors and what are the processes involved? This is addressed in the next sections.

AUDIT AND CERTIFICATION PROCESS

The **Requirements for bodies providing audit and certification of candidate trustworthy digital repositories** specified the way in which the audits were to be carried out. This defines the process and also the people. It is meant primarily for those setting up and managing the organization performing the auditing and certification of digital repositories.

It should also be of use to those who work in or are responsible for digital repositories seeking objective measurement of the trustworthiness of their repository and wishing to understand the processes involved. The document addresses issues arising from applying good audit practice to auditing and certifying whether and to what extent digital repositories can be trusted to look after digitally encoded information for the long-term, or at least for the period of their custodianship of that digitally encoded information.

It covers principles needed to inspire confidence that third party certification of the management of the digital repository has been performed with

- impartiality,
- competence,
- responsibility,
- openness,
- confidentiality, and
- responsiveness to complaints

The document specifies the ways of ensuring that the body providing such third party certification can inspire this confidence. It does this by building on the more general specifications of standards [14]-[16] which provide the framework and principles which must underlie an ISO audit. It must be understood that this is not a yes/no certificate which lasts forever. Instead the aim is to define a continuing process for improvement, with an audit producing certification subject to an improvement plan which is then followed up with a surveillance audit and then a subsequent re-certification audit – and then the cycle repeats.

Who are the auditors?

It is not possible to define with any precision what an auditor must know. However the general principle for accrediting an auditor is laid out. The auditor must have undertaken an accredited training course, must have appropriate expertise and must have taken part in audits with an existing group of auditors.

Who audits the auditors?

To bootstrap the process an initial body of auditors is defined – based, as we believe is reasonable, on the membership of the body which wrote the metrics document. This is called the Primary TDR Authorisation Body (PTAB). This body accredits training courses, undertakes the audits which the first batch of candidate auditors take part in, and accredits auditors. It also accredits national authorization bodies which will accredit auditors within individual countries, allowing for the creation of an international network of accreditation bodies.

Business Model, Scale and Scalability

The process which is defined should allow the number of auditors to be scaled up according to the demand. In particular the assumption is that there may be a significant demand from the commercial world. Indeed the business model must be that the audit process is self-funding yet the public sector including the cultural sector will not be willing or able to pay significant amounts. A healthy business sector demand would allow us to have a sustainable organisation.

Repository managers themselves may be unwilling to undergo an audit but the assumption behind the initial business plan is that those funding the repository would be willing (perhaps anxious) to have such an audit for their repository – if only to check the claims of those they fund. Moreover digital preservation is still a relatively new profession and not all problems are solved. The fact that there is not a “right” answer to every problem is another reason that external judgments will be sought.

EUROPEAN FRAMEWORK

In July 2010 the EU convened a meeting in which a Memorandum of Understanding was signed by

- David Giaretta in his capacity as chair of the CCSDS/ISO Repository Audit and Certification Working Group (RAC),
- Henk Harmsen in his capacity as Chair of the Data Seal of Approval (DSA) Board and

- Christian Keitel in his capacity as Chair of the DIN Working Group "Trustworthy Archives – Certification"

The MoU was to define a European Framework for Audit and certification of Digital Repositories. The framework consists of a sequence of three levels, in increasing trustworthiness:

- BASIC CERTIFICATION is granted to repositories which obtain DSA certification;
- EXTENDED CERTIFICATION is granted to Basic Certification repositories which in addition perform a structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644;
- FORMAL CERTIFICATION is granted to repositories which in addition to Basic Certification obtain full external audit and certification based on ISO 16363 or DIN 31644.

The advantage of this framework is that it provides an easy route up to full ISO certification.

TEST AUDITS

To support the European Framework a number of test audits were planned and undertaken. Three European repositories volunteered to be audited. The expenses of the auditors and the effort of the repositories to prepare for and to take part in the audits, was funded by the EU via the APARSEN project [17]. The three repositories were part of the UKDA[18], CINES [19] and DANS [20].

In addition three repositories in the USA, namely NSSDC [21], SEDAC [22] and the Kentucky Department for Libraries and Archives [23] volunteered to take part – note that they could not be funded by the EU and so they contributed their effort freely.

The audits were undertaken by PTAB members in June and July 2011. A normal audit would be undertaken by two auditors; these test audits were undertaken by larger groups as explained below. At the time of writing the final reports are being prepared.

It must be understood that the test audits had several aims:

- Support the creation of the European Framework – providing evidence of the usefulness of such audits
- Identify metrics which were poorly explained and difficult for repositories to understand. This information was used in the final (small) update of the metrics document
- Verify that the PTAB members had a common understanding and interpretation of the metrics, and judged evidence in a broadly similar way. It also helped to draft some of the audit operating procedures which would be used in real audits

The last bullet was particularly important because it gave confidence that if a repository were to be audited by two separate pairs of auditors then broadly similar evaluations would be produced. For this reason the test audits had to be undertaken by groups of PTAB members, to allow exchange of ideas and interpretations when faced with real evidence. In all cases recommendations for improvement were drawn up.

NEXT STEPS

The PTAB legal entity is being created, as required by [13]. Available training sessions are being reviewed to see which can be accredited for candidate auditors; an APARSEN Summer School, which will be tailored to auditor needs, is being penciled in also. The European Framework will be promoted and the commercial world will be targeted for information about the audit system. A dedicated web site will be set up.

CONCLUSION

The demand for a way to judge the ability of repositories to preserve digital information has been around for, in digital timescales, a long time. We believe that we have a way to provide this service in a way which fits into the mainstream ISO audit and certification mechanism and which is based on OAIS and the evidence it demands. We further believe that the process and organization we are setting up is scalable and can have the confidence of those who rely on such repositories.

REFERENCES

[1] Garrett J, Waters D (eds) (1996). Preserving Digital Information, Report of the Task Force on Archiving of Digital Information commissioned by The Commission on Preservation and Access and The Research Libraries Group. Available from http://www.ifla.org/documents/libraries/net/tfadi-fr.pdf

[2] Reference Model for an Open Archival System (ISO 14721:2002), http://public.ccsds.org/publications/archive/650x0b1.pdf or later version. At the time of writing the revised version is available at http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf or elsewhere on the CCSDS web site http://www.ccsds.org

[3] National Science Foundation Cyberinfrastructure Council (NSF, 2007), Cyberinfrastructure Vision for 21st Century Discovery. Retrieved from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

[4] TRAC (2007), Trustworthy Repositories Audit & Certification: Criteria and Checklist. Available from http://www.crl.edu/PDF/trac.pdf

[5] RLG-OCLC (2002) Report on Trusted Digital Repositories: Attributes and Responsibilities. Available from http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf

[6] nestor Working Group Trusted Repositories – Certification, (2006), Catalogue of Criteria for Trusted Digital Repositories. English version retrieved from http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

[7] CRL,(2007) Retrieved from http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92

[8] Giaretta D (2008) Comparison of OAIS and the Chicago Meeting 10 points. Available from http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/ComparisonOaisAndChicago10Points

[9] Ross S., Bütikofer N, McHugh A (2006), DCC Comments on RLG/NARA Audit and Certification Checklist. Available from http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/Ross_McHugh_Buetikofer_comments_RLGNARA_AUDIT_ver2.pdf

[10] Dale R (2007) Mapping of Audit & Certification Criteria for CRL Meeting (15-16 January 2007). Available from http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/TRAC-Nestor-DCC-criteria_mapping.doc

[11] Repository Audit and Certification Working Group http://wiki.digitalrepositoryauditandcertification.org

[12] Audit and Certification of Trustworthy Digital Repositories – review copy available from http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206520R1/Attachments/652x0r1.pdf. Final version should be available free from the CCSDS site http://www.ccsds.org

[13] Requirements for bodies providing audit and certification of trustworthy digital repositories http://wiki.digitalrepositoryauditandcertification.org or from the CCSDS web site http://www.ccsds.org

[14] ISO 19011:2002, Guidelines for quality and/or environmental management systems auditing

[15] ISO/IEC 17021:2006, Conformity assessment — Requirements for bodies providing audit and certification of management systems

[16] ISO/IEC 17000:2004, Conformity assessment — Vocabulary and general principles

[17] APARSEN project see http://www.aparsen.eu

[18] UK Data Arhive see http://www.data-archive.ac.uk/

[19] Centre Informatique National de l'Enseignement Supérieur see http://www.cines.fr

[20] Data Archiving and Networked Services – see http://www.dans.knaw.nl/en

[21] National Space Science Data Center – see http://nssdc.gsfc.nasa.gov/

[22] Socioeconomic Data and Applications Center – see http://sedac.ciesin.org/

[23] See http://kdla.ky.gov