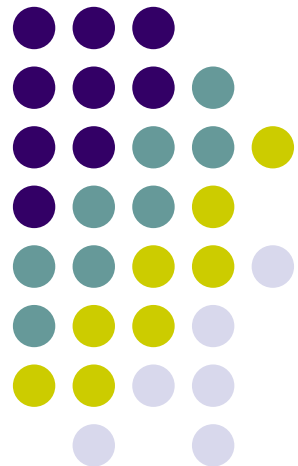
International Audit and Certification of Digital Repositories

PV 2009

David Giaretta



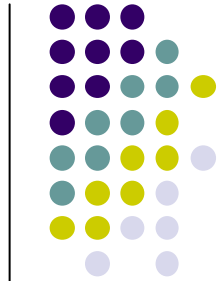
CCSDS.ORG
The Consultative Committee for Space Data Systems

CASPAR

DCC

PARSE insight

Digital Preservation…



- Easy to do…
- …as long as you can provide money forever
- Easy to test claims about repositories…
- …as long as you live a long time

Demand for a certification process

The Preserving Digital Information report of the Task Force on Archiving of Digital Information (Garrett & Waters, 1996) declared:

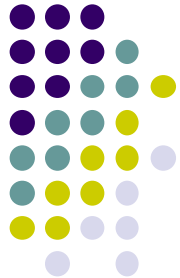
- **a critical component of digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating, and providing access to digital collections.**
- **a process of certification for digital archives is needed to create an overall climate of trust 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.

OAIS

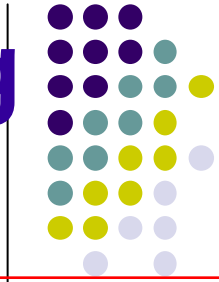
- Reference Model for Open Archival Information System (OAIS) provides an approach
  - Provides vocabulary – widely applicable
  - Conformance defined as mandatory responsibilities plus Information Model
  - Does not cover finance etc
- OAIS approach to digital preservation:
  - covers all types of digitally encoded information
  - provides a way to **test** whether preservation is successful
  - does not require seeing into the future
  - does require transparency
    - but does not require “open access”
  - does not cover social and organisational aspects
- OAIS does provide a good basis for certification

Key OAIS Concepts



- Claiming “This is being preserved” is untestable
  - Essentially meaningless
    - Except “BIT PRESERVATION”
- How can we make it testable?
  - Claim to be able to continue to“do something” with it
    - Understand/use
      - Need Representation Information
- Still meaningless…
  - Things are too interrelated
    - Representation Information potentially unlimited
  - Designated Community
- Many other concepts identified
  - Checklist – not just blanket term of “metadata”

Information is the important thing



- What information?
  - Documents……
  - Data…….
- Original bits?
- Look and feel?
- Behaviour?
- Performance?
- Explicit/ Implicit/ Tacit

Information:

Any type of knowledge that can be exchanged. In an exchange, it is represented by data.

Long Term is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community. Long Term may extend indefinitely.

Ensure that the information to be preserved is Independently Understandable to (and usable by) the Designated Community.

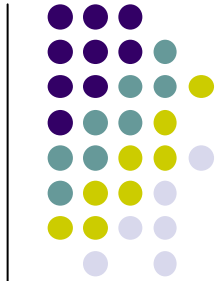
Issues of transferring info to future custodians

How can we ensure that the information trapped in the “bits” remains **understandable** despite all these changes?

- Things change:
  - Software
  - Hardware
  - Environment
    - E.g. Network links to related information
  - People
    - What is “common knowledge”
  - Organisations and systems
- Chain of preservation
  - Only as strong as its weakest link

How can current custodian prepare for or even be aware of these changes?

RLG/NARA work



- Part of the OAIS Roadmap
- Delegated to RLG and NARA to carry forward
- Plan to bring this come back to CCSDS for standardisation process.
- Based on
  - OAIS – technologies
  - TDR report – Finance, Organisational

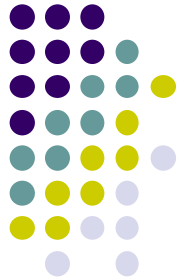
TRAC related work

- Trusted Digital Repositories: Attributes and Responsibilities from RLG and OCLC http://www.rlg.org/legacy/longterm/repositories.pdf
- Comments on the DRAFT RLG/NARA Audit and Certification Checklist (the "DCC document") http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/Ross_McHugh_Buetikofer_comments_RLGNARA_AUDIT_ver2.pdf
- Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) also available from http://www.crl.edu/PDF/trac.pdf
  - the earlier draft was: RLG/NARA Audit Checklist: http://www.rlg.org/en/page.php?Page_ID=20769
- TRAC-Nestor-DCC-criteria_mapping.doc: Crosswalk file between TRAC, Nestor and DCC work, which was completed by Robin Dale as a part of the Center for Research Libraries project http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/TRAC-Nestor-DCC-criteria_mapping.doc

Other related work

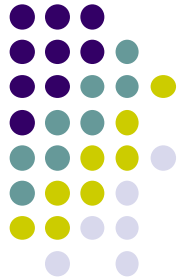
- English version of the nestor criteria catalogue: http://edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf
- OECD Guidelines for the Security of Information Systems and Networks http://www.oecd.org/dataoecd/16/22/15582260.pdf
- The outcome of the related Chicago meeting is available:
  - Notes from a related meeting in Chicago 15-16 Jan 2007 http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/Chicago_meeting.doc
- DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) - see http://www.repositoryaudit.eu/
- Joint meeting of “Audit and Certification Forum” in Berlin 27 Nov 2007 agreed to use RAC as a clearing house after private discussions within the various groups (nestor, DRAMBORA, CRL etc)

Repository Audit and Certification Working group



- Created as CCSDS “Birds of a Feather” (BoF) group in CCSDS
- Now an official CCSDS Working Group
- Open virtual meetings, notes and documents:
  - http://www.digitalrepositoryauditandcertification.org

Background



- Working Group of CCSDS
  - Charter of WG agreed at meeting January 2007
  - Goal: Obtain ISO approval of a standard that establishes the criteria that a repository/archive must meet to be designated an ISO Trusted Digital Repository
    - on which a full audit and certification of digital repositories can be based
  - Following route of OAIS
    - CCSDS is the “working arm” of TC20/SC13 of ISO
  - Based on TRAC document

RAC Charter

- Goal 1: Obtain ISO approval of a standard that establishes the criteria that a repository/archive must meet to be designated an ISO Trusted Digital Repository.

1. Review the existing work on audit and certification criteria for digital repositories, such as that from the RLG/NARA working group and the NESTOR project. These two documents are broadly similar, and both are based on the OAIS Reference Model.
2. Prepare a draft (or adopt one of the above documents) and submit to ISO as a Committee Draft to get the ISO process going.
3. Analyse the consistency of those works with the OAIS Reference Model (ISO 14721) and follow on standards such as PAIMAS and the forthcoming PAIS.
4. Review existing audit and certification standards such as ISO 9000 and ISO 27000, and the requirements on such standards for supporting an accreditation and certification programme to obtain guidance on the form of this standard. Neither of these two standards audit the preservation of the encoded information, hence the need for a new standard.

Participation

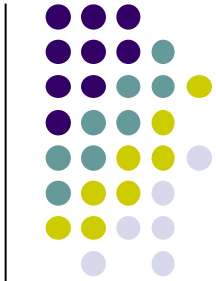
- UK
  - STFC
  - HATII, U Glasgow
  - Digital Curation Centre, UK
- European Space Agency
- France
  - CNES
- Netherlands
  - KB National Library of the Netherlands
- Germany
  - nestor
- USA
  - NASA/GSFC/NSSDC
  - ICPSR
  - Smithsonian Institution Archives
  - California Digital Library
  - Center for Research Libraries
  - National Archives and Records Administration
  - Columbia University
  - U Maryland
  - UNC
- Brazil
  - Instituto Nacional de Pesquisas Espaciais INPE

Mailing list

- USA 40
- South Africa 8
- Australia 6
- China 3
- Israel 3
- Canada 1
- India 1

- UK 20
- Germany 6
- France 5
- ESA 4
- Netherlands 2
- Italy 2
- Spain 1
- Ireland 1
- Czech Republic 1
- Estonia 1

Current status

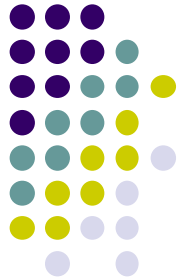


- An open process with publicly accessible wiki:
  - www.digitalrepositoryauditandcertification.org
- Weekly online discussions metrics document and other documentation required
  - Notes recorded on wiki with all working documents
- Agreed as the “clearing house” for the private discussions of the other groups in this area
- Major progress at face-to-face meeting in Washington in February

Current status

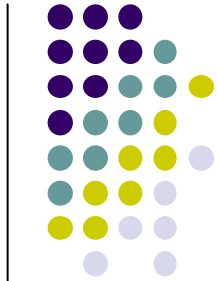
- Now converging fast to agreed documents
  - Audit and Certification of Trustworthy Digital Repositories
    - **IN REVIEW PROCESS**
  - Requirements for Bodies Providing Audit and Certification of Digital Repositories
    - Almost ready for review

Issues encountered



- Science data – continuing understandability
- Authenticity and Information Properties (new OAIS term)
- Links to ensure consistency with OAIS update
  - Note that the OAIS update is now complete and is in the ISO review process
- Level of detail of metrics
  - Lessons from others ISO standards e.g. Info. security

Structure and approach



- Section A: Organisational Infrastructure
- Section B: Digital Object Management
- Section C: Infrastructure and Security Risk Management
- Metrics and their structure:
  - Statement of requirement
  - Supporting text
  - Examples of Ways the Repository can Demonstrate it is Meeting this Requirement
  - Discussion

C1 Technical infrastructure risk management (RM)

C1.1 Repository identifies and manages the risks to its preservation operations and goals associated with system infrastructure.

Supporting Text

The repository must conduct or contract assessments of the risks related to hardware and software infrastructure, and operational procedures. The repository must provide mechanisms that minimize risk from dependencies on proprietary or obsolete system infrastructure. The repository must provide mechanisms to minimize risk from operational error. This is necessary to ensure a secure and trustworthy infrastructure.

Examples of Ways the Repository can Demonstrate it is Meeting this Requirement

Infrastructure inventory of system components; periodic technology assessments; estimates of system component lifetime; export of authentic records to an independent system; use of strongly community supported software .e.g., Apache, iRODS, Fedora); re-creation of archives from backups.

Discussion

The degree of support required relates to the criticality of the subsystem(s) involved in long- term preservation. The repository should maintain a system that is scalable (e.g. able to handle anticipated future volumes of both bytes and files) without a major disruption of the system. The repository should maintain a system that is evolvable. That is, the system should be designed in such a way that major components of the system can be replaced with newer technologies without major disruption of the system as a whole. The repository system should be extensible. That is, the system should be designed to accommodate future formats (media and files) without major disruption of the system as a whole. The repository should be able to export its holdings to a future custodian. The repository should be able to re-create the archives after an operational error that overwrites or deletes digital holdings.

Key Issues (1)

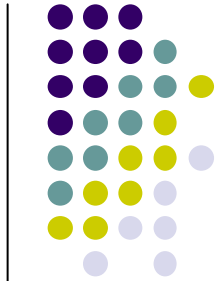
- How to get from a checklist to an international accreditation/ certification system?
- The initial auditors
- Qualification for auditors
- International set-up

Key Issues (2)

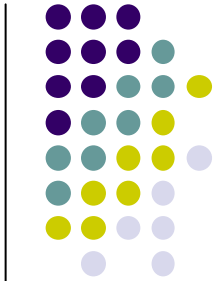
- Evidence – short term
- Evidence – long term
  - The real crunch!
- Risk capture
  - e.g. DRAMBORA tool
- Quantification
  - The marking system
- Levels of audit?
  - External review
  - Internal maturity
- Legal issues

The Market

- Transparency
- Trustable?
  - certified by whom?
  - to what level?
  - what evidence?
  - with what granularity?
    - for what Designated Community
      - relevant/sensible?
- What cost?
  - Self-sustaining?



Next steps



- Standards should be approved by Q2 2010
- In early 2010 set up initial audit committee
- Seek support of funders to do initial audits on a selection of their archives
  - Provide evidence of usefulness of audits to funders
  - Decide pricing structure
  - Ensure consistency of audits results
- Commitment of audit of repositories of funders
- Set up national committees
- Auditor training
- Start auditing commercial repositories

END