Trusted Repository Certification - Challenges of Scale

MSST2010 - Symposium on Massive Storage Systems and Technologies

David Giaretta

May 4-52010, Lake Tahoe, Nevada, USA

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

What is wanted

- By Repositories:
  - Comfortable
  - Low cost
  - Low trouble
  - Something to confirm they are going a good job
- By Funders
  - How to tell – independently – that money is been spent well
    - Otherwise risk money being wasted and data lost
  - International standard – preferably ISO

Challenges of Scale

- Not enough experience (by anyone) of long term preservation of massive amounts of data
- How can audit/certification provide any kind of judgement?

Digital Preservation

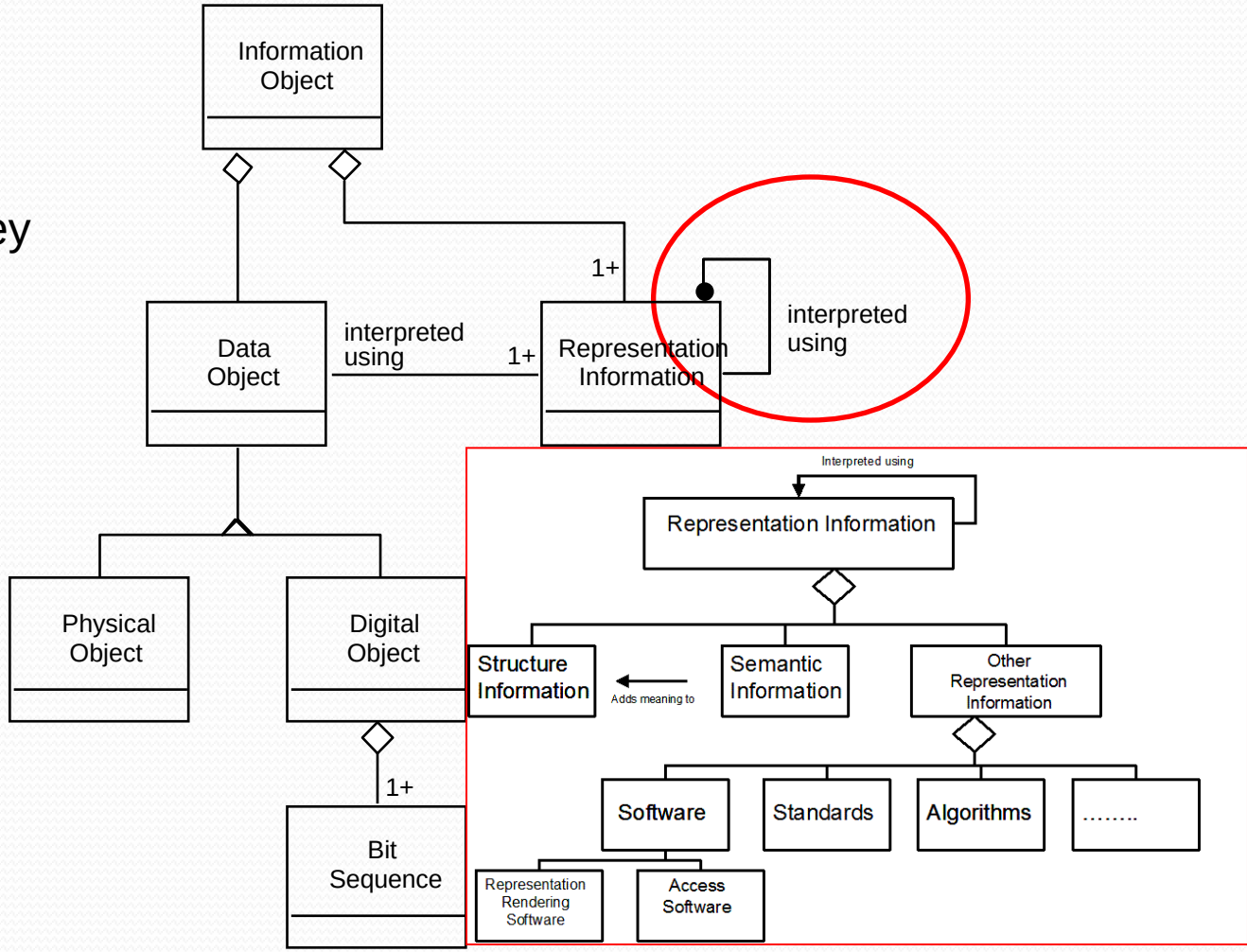
- Ensure that digitally encoded information are understandable and usable over the long term
  - Long term could start at just a few years
- Easy to make claims
  - Difficult to provide proof
- Reference Model for Open Archival Information System (ISO 14721)
  - The basic standard for work in digital pres.
  - Defines terminology and compliance criteria

Information Model & Representation Information

The Information Model is key

Recursion ends at KNOWLEDGEBASE of the DESIGNATED COMMUNITY

(this knowledge will change over time and region)



Repository Audit and Certification Working group

- Closely related to OAIS Reference Model
  - Certification was identified as a follow-on standard
  - Following route of OAIS
    - CCSDS is the "working arm" of TC20/SC13 of ISO
  - TRAC work provided the initial draft
- CCSDS Working Group
- **Open** virtual meetings, notes and documents:
  - **http://www.digitalrepositoryauditandcertification.org**
  - **http://www.digitalrepositoryauditandcertification.org**

Metrics

- Available from *http://wiki.digitalrepositoryauditandcertification.org/pub/Main/MetricsRidResolution/652xor1candidate-update-typoscorrected.doc*

- 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

Level of detail

- Impossible to anticipate all possibilities
- Other standards (e.g. ISO 2700x security standards) are quite brief
- Should be regarded as only a “guide” for Audit Team
- Fundamentally depends on audit experience
  - “Requirements for Bodies providing Audit and Certification” defines how the audit/certification organisation operates to ensure:
    - certification bodies operate management system certification in a competent, consistent and impartial manner
  - facilitating the recognition of such bodies and the acceptance of their certifications on a national and international basis.

What would Certification look like?

- Not a simple statement that “Yes this repository is perfect”!
- Should be regarded as part of a process of improvement
  - Audit/certification provides information on which an organization can act to improve its performance
  - Cycle of certification/ surveillance audit/ re-certification
- May be possible to define maturity levels

Possible European Framework

- BRONZE level:
  - Data Seal of Approval
    - Monitored self-audit
    - Published evidence for a small number of criteria
- SILVER level:
  - DSA plus self audit using RAC metrics
    - Published evidence
- GOLD level
  - Full ISO audit

Challenges of Scale

- Not enough experience (by anyone) of long term preservation of massive amounts of data
- How can audit/certification provide any judgement?
- How can any improvements be recommended?

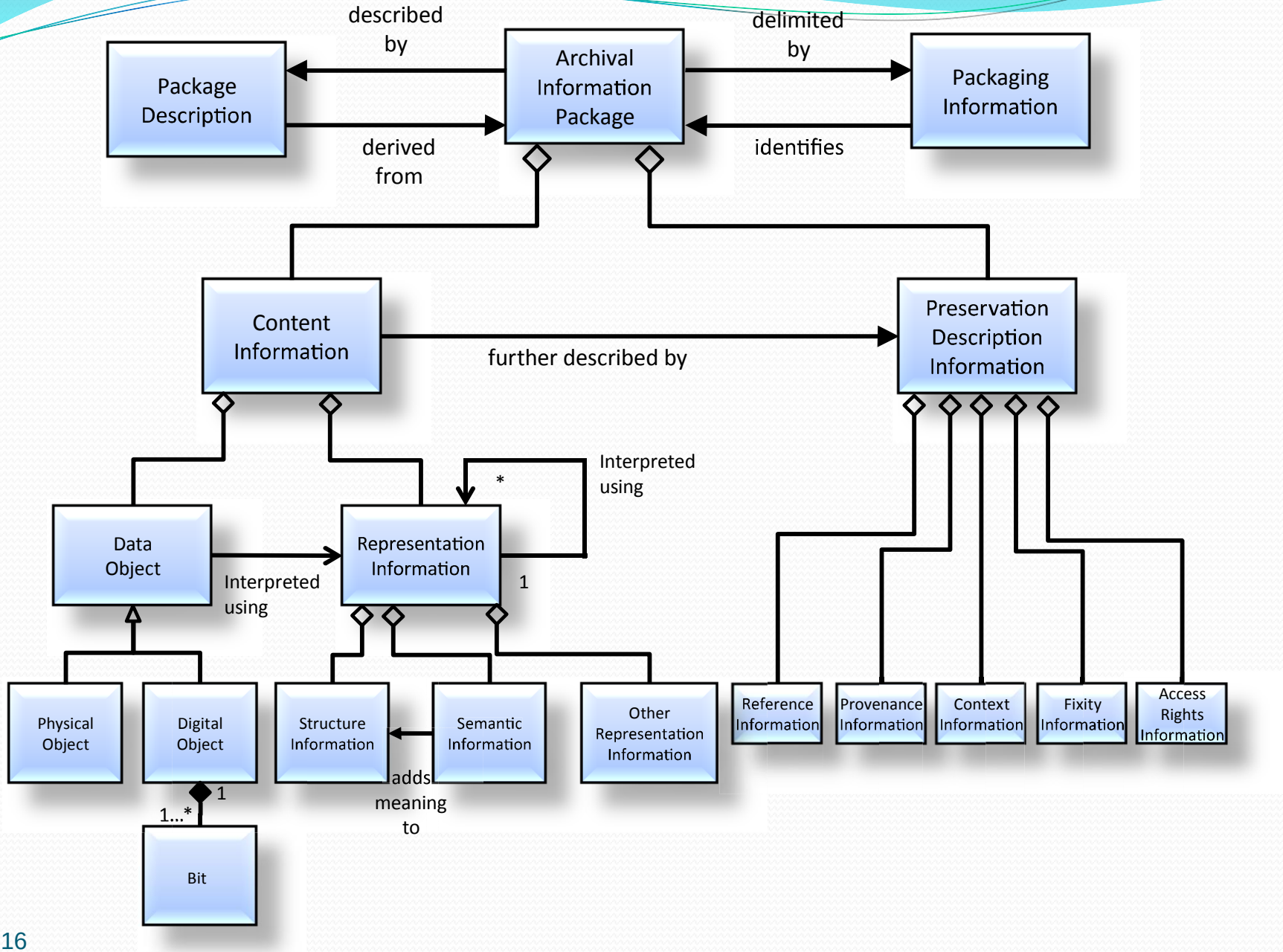
What can be done?

- The audit team can ask a number of basic questions e.g.
  - Are the bits safe?
  - Are the data understandable/usable by the **Designated Community**?
  - Is **authenticity** safeguarded (evidence based)
    - E.g. Are the bits really what they are claimed to be?
  - Can the digital holdings be handed over to another repository if/when necessary?
- The repository must try to provide evidence
  - Why do they think people should trust them?
- Learning process – over several audit cycles
  - Comes into focus when repositories make claims about digital preservation

Links

- **Repository Audit and Certification working group**
  http://wiki.digitalrepositoryauditandcertification.org
- **ISO submission of audit Metrics**
  http://wiki.digitalrepositoryauditandcertification.org/pub/Main/MetricsRidResolution/652x0r1candidate-update-typoscorrected.doc
- **OAIS Reference Model**
  - ***Original version available from***
    http://public.ccsds.org/publications/archive/650x0b1.pdf
  - **Updated version is available from**
    http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/OAIS-after-CCSDS-review.pdf

END


described by
Package Description
derived from
Archival Information Package
delimited by
identifies
Packaging Information
Content Information
further described by
Preservation Description Information
Interpreted using
*
Data Object
Interpreted using
Representation Information
1
Physical Object
Digital Object
1
1...*
Bit
Structure Information
Semantic Information
adds meaning to
Other Representation Information
Reference Information
Provenance Information
Context Information
Fixity Information
Access Rights Information

The future

- Finalisation of "Requirements for Bodies providing Audit and Certification" document
  - Expect by end of May
- Conduct a number of test audits in APA
  - NOTE: these are tests of the document rather than testing the repositories
    - Do two separate auditors reach the same judgment of a repository?
      - Necessary (but not sufficient) condition to ensure clarity of metrics and common understanding
- ISO procedures
  - Expect end of ISO review by end of year
- Plans for the international accreditation and certification process will be completed during the ISO review

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”

Issues of transferring info to future custodians

- 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 weak link

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

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

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.

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)

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