

Trustworthiness of Preservation Systems

David Minor

UC San Diego Library

Why does anyone care?

- We all want to trust systems
- Preservation systems more than most
 - Long-term-ness
 - Often not-us-ness
- How do we say we can?
- What does trustworthy mean?

Why do you care?

- You're running a repository
- You're using a 3rd party repository

General principles

- Trust, verify, and clarify – an iterative process
 - Institutional commitment, intent for collections, objects
 - Infrastructure demands
 - Technical system and staffing capabilities
 - Sustainability (funding, technology, collaboration)
 - And if all else fails? “Fail-safe” partners?

General principles, con't

- Identify and communicate context
 - The “why” questions
 - The “who cares” questions
- Identify and communicate risks to content
 - The “what if” questions

General principles, con't

- Future planning
 - Identify best practices
 - Map best practices against current system
 - Identify gaps in practice, policy, procedure, functionality
 - Map best practices against local capabilities/goals
 - Identify opportunities for future expansion and needs associated with them

Three levels of auditing

European Framework for Audit and Certification of Digital Repositories:

- "Basic certification" is a simple self-assessment
- "Extended certification" represents a plausibility-checked self-assessment
- "Formal certification" is an audit by external experts.

<http://www.trusteddigitalrepository.eu>

Major auditing frameworks

- Data Seal of Approval (Basic)
- Nestor (Extended)
- TRAC/ISO 16363 (Formal)
- DRAMBORA (Range)



Data Seal of Approval (DSA)



Data Seal of Approval (DSA)

- Long-term archiving of *research data*
- Guarantee the **durability** of archived data
- 16 requirements
- Stage 1: Self assessment via DSA online tool
- Stage 2: Archive/Repository assessed by peers – review based on *trust*
- Stage 3: Awarded DSA



Data Seal of Approval (DSA)

The 16 Data Seal of Approval guidelines are based on five criteria:

- The data can be found on the Internet
- The data are accessible (clear rights and licenses)
- The data are in a usable format
- The data are reliable
- The data are identified in a unique and persistent way so that they can be referred to



Data Seal of Approval (DSA)

Approximately 35 institutions have passed the test and acquired the seal.



<http://datasealofapproval.org/en/assessment/>



nestor seal for trustworthy
digital archives



nestor



- More elaborate than a simple self-assessment
- Language in the documentation references the EU designation: is intended to be an “Extended certification” via a plausibility-checked self-audit
- Based on DIN 31644, “Criteria for trustworthy digital archives”

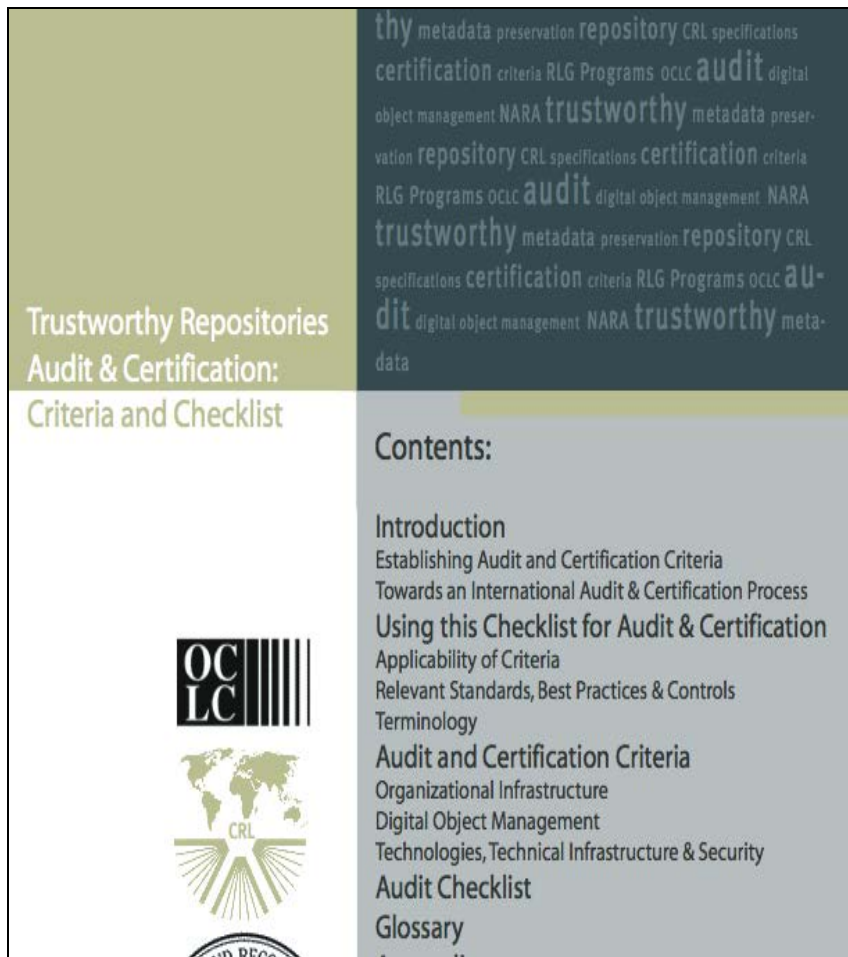
nestor process

- Initial self-audit completed
- Reviewed by an external reviewer, changes or edits made as needed
- Reviewed by a third external reviewer



Trustworthy Repositories Audit & Certification (TRAC) / ISO 16363

Trustworthy Repositories: Audit & Certification, 2007 (TRAC)



- Criteria checklist for evaluating trustworthiness of repositories & archives
 - Derived from TDR, OAIS and other standards
- Helpful for ...
 - Digital repository planning
 - Digital repository assessment/audit/evaluation/gap analysis/weaknesses
 - Evaluating third party service provider capabilities
- *Evidence-based assessment*
- Focusing on Designated Community Needs

Assessment of...

1. Organizational infrastructure

- Governance, structure, staffing, policies, procedures, finances, contracts, licenses, etc.

2. Digital object management

- Acquisition, creation of archival package, preservation planning, archival storage, information [metadata] management, access management [designated community]

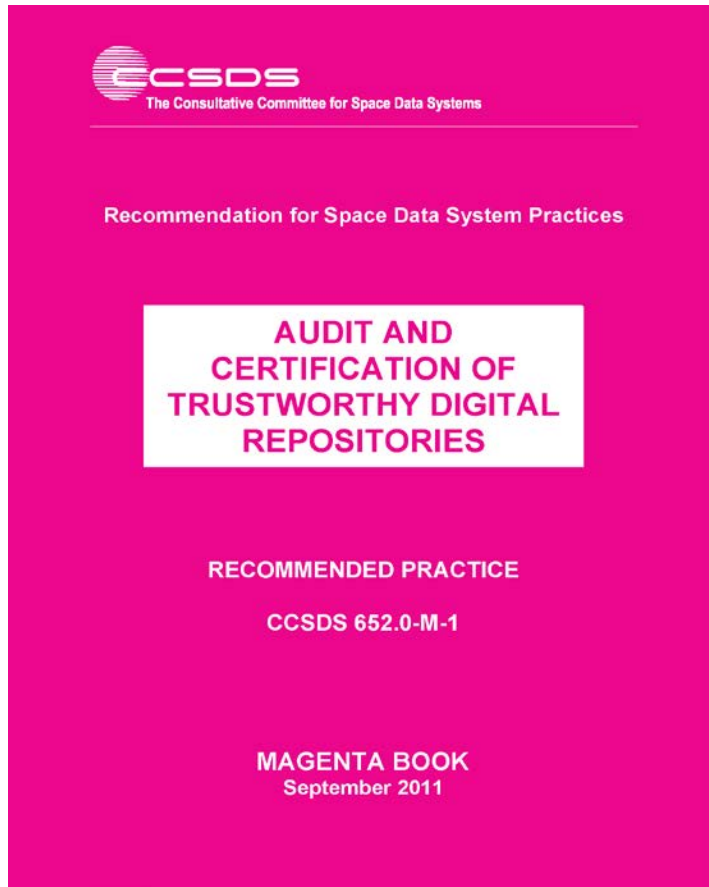
3. Technologies, technical infrastructure, & security

- System infrastructure, appropriate technologies, security

TRAC Has Been Utilized By...

- Organizations doing gap analysis
- Organizations building repositories
- Center for Research Libraries to certify repositories and/or services
 - Portico
 - HathiTrust
 - Chronopolis
 - Scholar's Portal
- But others wanted it to become an international standard...

Four + years later... ISO 16363: 2012



- An international standard is born
- Part of the OAIS family of standards
- Very similar in nature to TRAC, content-wise
- 3 Sections, 109 criteria
- For formal audit, should be done with auditors conforming to ISO/DIS 16919

ISO 16363 is Utilized By...

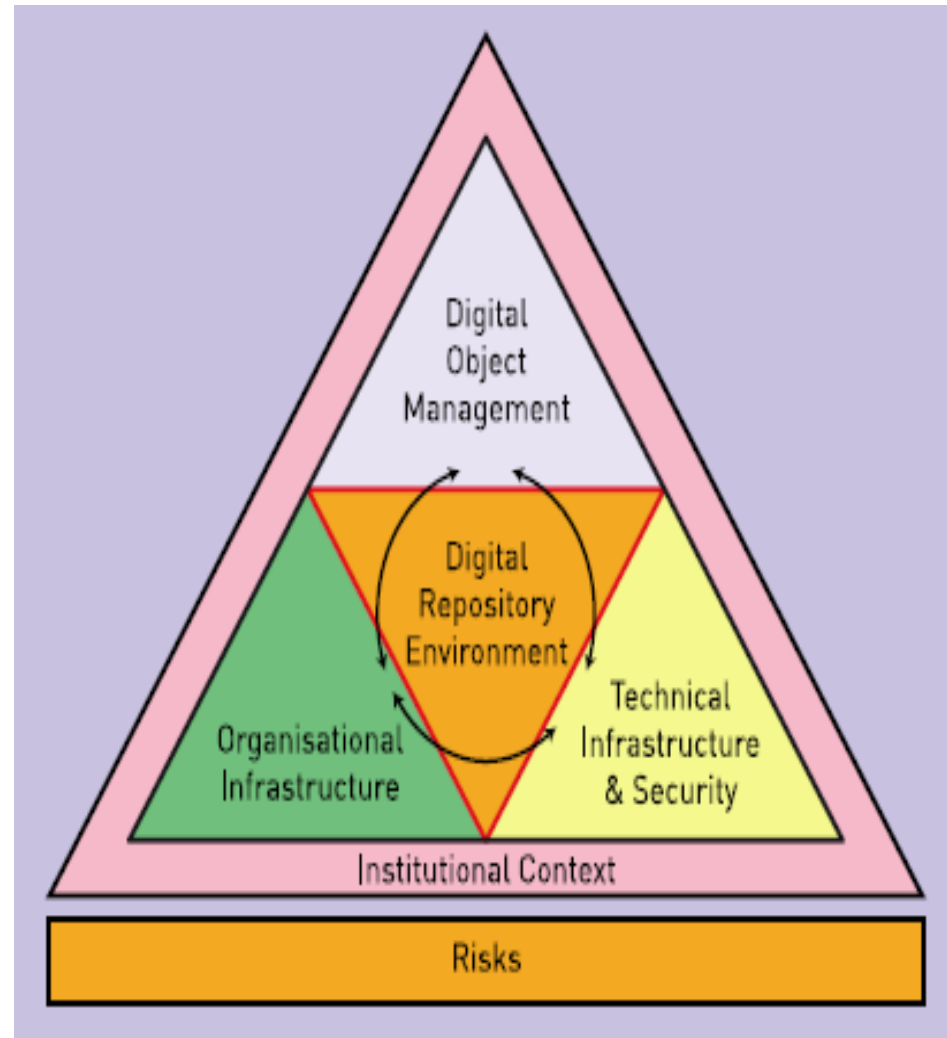
- **No one!**
- Officially, must be completed by ISO/DIS 16919, still in the standards approval stage, *until this can be formalized, NO ISO 16363 formal audit can be conducted nor can certifications be issued.*
 - *As of 7-8-2014, it is now out for ballot for 2 months*
 - Also, only 1 organization – Primary Trustworthy Auditing Body or PTAB – has been set up
 - Primarily comprising people who worked on ISO 16363
 - Based in Europe
- Unofficially, it can be used as the same gap analysis / risk management / weakness identification tool as TRAC
 - Could better prepare you for international certification if that is the end goal you seek
 - ISO 16363 Self Assessment Tool available

... a couple other things

DRAMBORA

Digital Repository Assessment Method Based on Risk Assessment

- Linked to TRAC development
- Initial focus on existing repositories and data already entrusted
- Self & facilitated audit options
- “seeks to determine whether the repository has made every effort to avoid and contain risks that might impede its ability to receive, curate and provide access to authentic, and contextually, syntactically and semantically understandable digital information.”



DRAMBORA Stages

1. Identify organizational context
2. Document policy and regulatory framework
3. Identify activities, assets, and their owners
4. Identify risks
5. Assess risks
6. Manage risks

<http://www.repositoryaudit.eu>

Need to Evaluate Existing Tools or Systems?

Digital POWRR Tool Grid

Digital POWRR Tool Evaluation Grid	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Bit Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
ACE (Audit Control Environment)		x																	x		x			Free
AFF Open Source Computer Forensics Software							x															x		Free
Amazon S3	x	x			x	x									x	x	x	x		x	x		x	Varies
Archive-It	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x		x		x	x	Varies
**Archivematica	x	x	x		x	x	x	x	x	x	x		x	x					x			x	x	Free
BagIt Library		x			x	x				x				x										
BagIt Transfer Utilities																								Free
BitCurator	x	x				x	x	x		x		x												free
BWF MetaEdit		x						x	x													x		free
Carbonite	x	x													x	x		x		x	x			\$149 – \$599
Chronopolis	x	x		x	x	x	x	x	x	x	x			x	x	x	x	x	x	x	x		x	\$1,500 – \$2,200
Cinch		x	x	x	x		x																x	Free
ContextMiner																						x	x	Free

In the future

- How do these audit frameworks apply to distributed digital preservation environments?
- How strict are the questions?
- How flexible are the models underlying the audits (e.g. OAIS)?
- Are we driving toward a single standard?

Thank you!

David Minor

dminor@ucsd.edu