Trusted Repositories

**DPE/Planets/CASPAR/nestor
Joint Training Event:
Starting out: Preserving Digital Objects -
Principles and Practice**

**October, 13th- 17th 2008
Prague, Czech Republic**

**Stefan Strathmann
Göttingen State and University Library (SUB)**

Special thanks goes to Susanne Dobratz & Astrid Schoger from the nestor-
WG „Trusted Repositories – Certification"

Session Outline

10:00 – 11:00 Lecture

11:00 – 11:30 Discussion

11:30 – 11:45 Coffee Break

11:45 – 12:45 Group Work

12:45 – 13:30 Groups present their results and Summary discussion

Table of Contents

- General introduction

- Initiatives and activities

- Example: nestor Catalogue of Criteria for Trusted Digital Repositories
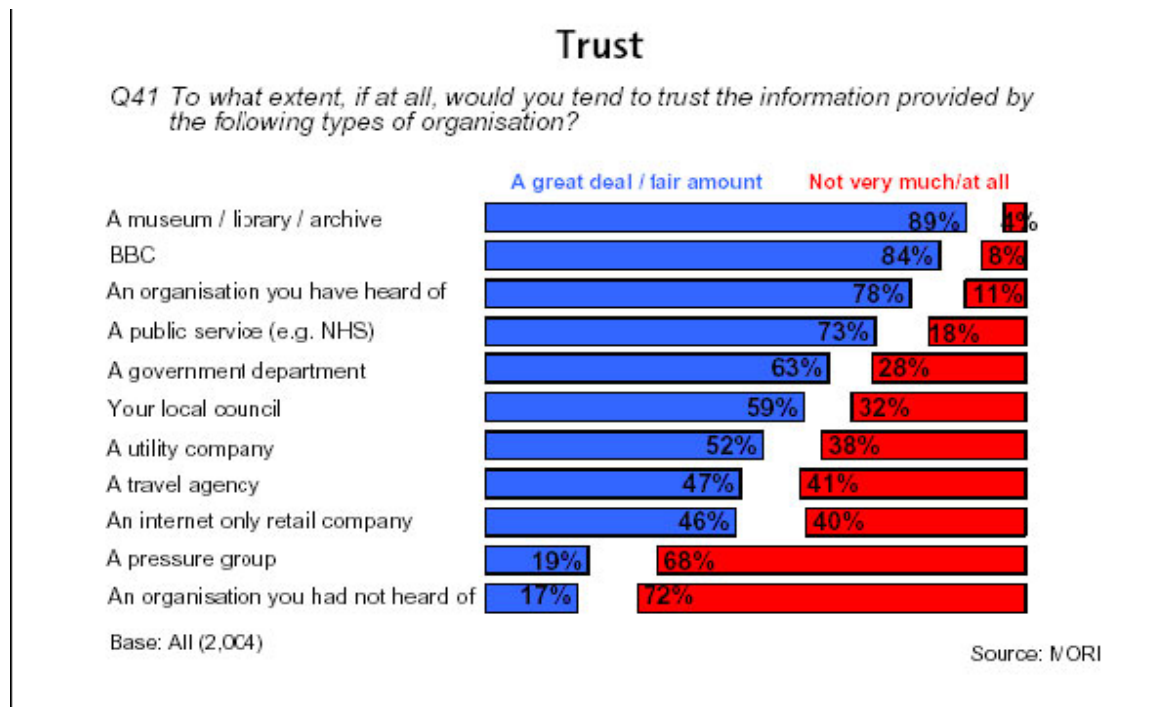
Trustworthy Digital Archives

"… repositories claiming to serve an archival function must be able to prove that they are who they say they are by meeting or exceeding the standards and criteria of an independently-administered program for archival certification .."

Task Force on Archiving Digital Information (1996): Preserving Digital Information, Commission on Preservation and Access, Washington D.C.

Digital Preservation & Trust

- Creation of digital information continues to accelerate!
- Digital information are valuable and vulnerable!
- Practical digital preservation/curation efforts are just starting.
- Who can guarantee the long-term availability, authenticity and integrity of digital information?
- Who is trustworthy? Which institutions, approaches and technologies can be trusted?

Trusted Information

Trust

Q41 To what extent, if at all, would you tend to trust the information provided by the following types of organisation?

	A great deal / fair amount	Not very much/at all
A museum / library / archive	89%	4%
BBC	84%	8%
An organisation you have heard of	78%	11%
A public service (e.g. NHS)	73%	18%
A government department	63%	28%
Your local council	59%	32%
A utility company	52%	38%
A travel agency	47%	41%
An internet only retail company	46%	40%
A pressure group	19%	68%
An organisation you had not heard of	17%	72%

Base: All (2,004)

Source: MORI

Who is interested in Trusted Repositories?

- General public, end user
- Information producer
- Archival Institutions: management, staff, responsible bodies
- Partner in a cooperative digital preservation (trusted repositories are the basis for cooperative digital preservation)

Challenges

- Broad variety of archives
- Different designated communities
- Variety of object types
- Different standards in use

- Authenticity of the objects
- Integrity of the objects

Authenticity

The object actually is what it claims to be!

- Complete authenticity / bit stream preservation
  - Detachment of the data from the original media
  - Transfer of the data into a homogeneous storage system
  - Refreshing
- Relative authenticity
  - Long-term preservation of the availability/usability (Look&Feel!) of digital objects
 - Regular migration may be required

Integrity

- Integrity refers to the completeness of the digital objects and to the exclusion of unintended modifications as defined in the preservation rules.

- Integrity is measured in terms of the characteristics of the digital object being preserved.

International Efforts – A Chronology

- 2002: RLG/OCLC Report: Trusted Repositories Attributes & Responsibilities
- 2002: Reference Model for an Open Archival Information System (OAIS)
- 2005: RLG/NARA: Audit Check-list for Repository Certification
- 2006: nestor: Catalogue of Criteria for Trusted Digital Repositories
- 2007: nestor/CLR/RLG/DPE/DCC: Core Requirements for Digital Archives
- 2007: DCC/DPE: Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)
- 2007: CRL/OCLC: Trustworthy Repositories Audit & Certification (TRAC): Criteria and Check-list

Trustworthy Repositories Audit & Certification (TRAC): Criteria and Check-list
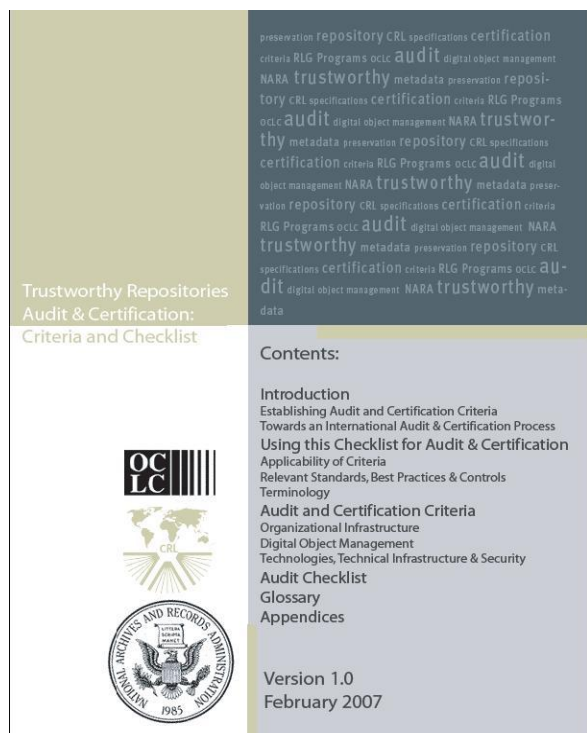
- Revised and expanded version of "The Audit Checklist for the Certification of Trusted Digital Repositories", originally developed by RLG-NARA
- Test audits conducted 2006/2007
- Provides Tools for the audit/assessment of digital repositories.
- Compiles documentation requirements.
- Drafts a certification process.
- Establishes methodologies for the determination of the sustainability of digital repositories.

http://www.crl.edu/content.asp?l1=13&l2=58&l3=162

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)

Digital Repository Audit Method

Based on Risk Assessment

DRAMBORA

Digital Curation Centre (DCC)

&

DigitalPreservationEurope (DPE)

Draft for Public Testing & Comment

Release: Version 1.0 (draft)
Date: 28 February 2007

D|C|C

digital preservation europe

- Based on risk-analyses
- Test audits in 2007 seq.
- Discussions in ISO Group TC46

http://www.repositoryaudit.eu/

10 Common Principles I

In January 2007 DCC, DPE, nestor and CRL agreed on 10 basic characteristics of digital preservation repositories:

The repository:

- Commits to continuing maintenance of digital objects for identified community/communities.

- Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.

- Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.

10 Common Principles II

- Has an effective and efficient policy framework.

- Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.

- Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.

- Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.

- Fulfills requisite dissemination requirements.

10 Common Principles III

- Has a strategic program for preservation planning and action.

- Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

The key premise underlying the core requirements is that for repositories of all types and sizes preservation activities must be scaled to the needs and means of the defined community or communities.

nestor

- nestor - Network of Expertise in Long-Term Storage of Digital Resources
- Duration: May 2003 – June 2006 and July 2006 – June 2009; continuation expected
- Funded by the German Ministry of Research and Education
- Partner: cultural heritage sector (libraries, archives, museums)
- Aim: coordination, information and communication - not archiving

nestor WG on Trusted Repositories Certification

- Broader group of members than nestor (+ World Data Center, Computer Scientists, Certification Specialists, …)

- Start in Dec. 2004

- Aim: a net of trustworthiness in which long-term digital archives can function in various environments (libraries, archives, museums…)

nestor WG on Trusted Repositories Certification

- Provide a coaching instrument to force a certain level for digital archives, ensure acknowledgements of recent standards
- Tight cooperation and permanent involvement of the communities
- Don't reinvent the wheel, but fit criteria into Germany's conditions
- nestor WG is transferred into the DIN NABD15 / mirror group to ISO TC46

Catalogue of Criteria
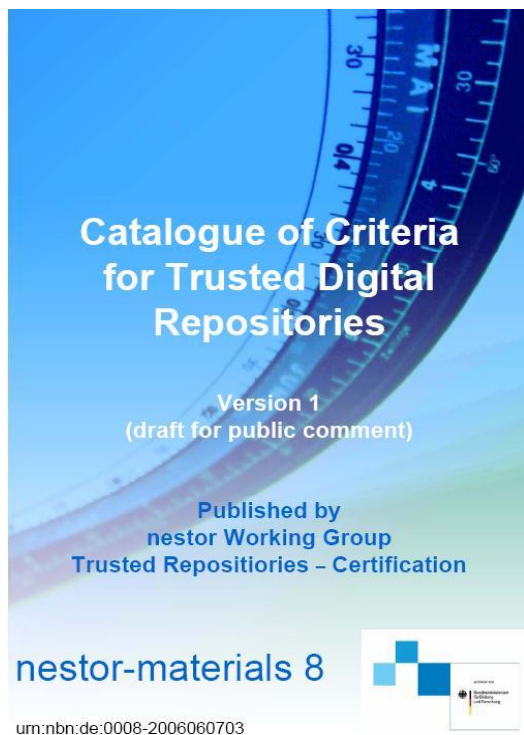
- nestor Catalogue of Criteria 2006
- At the moment: Preparation of second edition
- Self-assessment tool
- Target group: cultural heritage organizations, software developers, third party vendors, …
- Work will be continued in the DIN NABD15 / mirror group to ISO TC46

http://www.nbn-resolving.de?urn:nbn:de:0008-2006060703

Catalogue of Criteria for Trusted Digital Repositories

Version 1
(draft for public comment)

Published by
nestor Working Group
Trusted Repositiories – Certification

nestor-materials 8

urn:nbn:de:0008-2006060703

Catalogue of Criteria

- Formulates abstract criteria, enhanced with examples and explanations
- Focused on application in Germany, but orientated on international discussions and standards

Digital Long-term Archive

- … is defined as an organisation (consisting of both people and technical systems) that has assumed responsibility for the long-term preservation and long-term accessibility of digital objects, ensuring their usability by a specified target group, or 'designated community'.

- "Long-term" in this context means beyond technological changes (to hardware and software) and also any changes to this designated community.

- Once more, this definition of digital archive is based on the OAIS Reference Model.

Central Concepts of the Catalogue - Thrustworthiness

- Trustworthiness is the capacity of a system to operate in accordance with its objectives and specifications (that is, to do exactly what it claims to do).

- From an IT security perspective, the fundamental considerations are integrity, authenticity, confidentiality and availability.

- IT security is therefore an important prerequisite for trusted digital repositories.

Central Concepts of the Catalogue - Implementation

- Implementation (of the long-term archive and of single criteria) as a multi step process
  - 1. Conception
  - 2. Planning and Specification
  - 3. Realization and Implementation
  - 4. Evaluation
  - Because of permanent changes, these steps must be repeated if necessary (quality management)

Basic principles for application of criteria - Documentation

- The objectives, basic concept, specifications and implementation of the digital long-term repository should be well documented.

- The documentation can be used to evaluate the status of development both internally and externally. Early evaluation can serve to avoid errors caused by inappropriate implementation. Correct documentation of workflow also allows verification of any evaluatory conclusions.

- All quality and security standards must also be suitably documented.

Basic principles for application of criteria - Transparency

- Transparency is achieved by publishing appropriate parts of the documentation. It relates closely to trust as it permits interested parties to make a direct assessment of the quality of a digital repository.

- External transparency (users, partners, funders etc.) enables to gauge the degree of trustworthiness. Transparency afforded to producers and suppliers enables these groups to determine to whom they wish to entrust their digital objects.

- Internal transparency facilitates reflective self-assessment by the operators, managers and all other employees.

Basic principles for application of criteria - Adequacy

- The principle of adequacy derives from the fact that the conception of absolute standards is somewhat unfeasible; rather that evaluation is always based on the objectives and tasks of the individual digital repository concerned.

- The criteria have to be related to the context of each individual archiving task. Individual criteria may therefore prove irrelevant. Depending on the objectives and tasks of the digital repository, the required degree of compliance for a particular criterion may differ.

Basic principles for application of criteria - Measureability

1. In some cases - especially with regard to long-term issues - there are no objectively measurable characteristics.

2. In such cases we must rely on other indirect indicators that demonstrate the degree of trustworthiness. Again, transparency makes the indicators accessible for evaluation.

Composition of the Criteria

- The main criteria are on a very abstract level (because of the broad scope)
- They are enriched by subcriteria, detailed explanations, examples and references
- As basis for a common terminology the OAIS reference model was taken, where possible
- An audit checklist is provided together with the catalogue of criteria

Overview of Main Criteria I

A   Organizational Framework
1. Goals are defined
2. Adequate usage is guaranteed
3. Legal rules are observed
4. Adequate organization is chosen
5. Adequate quality management is conducted

Overview of Main Criteria II

B   Object Management

1.   Integrity of digital objects is ensured
2.   Authenticity of digital objects is ensured
3.   A preservation planning is implemented
4.   Transfers from producers are defined
5.   Archival storage is well defined
6.   Usage is well defined
7.   Data management guarantees the functionality of the repository

Overview of Main Criteria III

C   Infrastructure and Security

1. The IT infrastructure is adequate
2. The infrastructure ensures the protections of the repository and its digital objects

Structure of the criteria catalogue

- **Criterion**

- General explanation of the criterion

- Examples, comments, notes from different application areas, with no claim to exhaustiveness

- *Literature related to this criterion*

Example

A Organisational Framework

1 The repository has defined its goals.

- 1.1 Selection criteria
- 1.2 Responsibility for the long-term preservation of the information represented by the digital objects
- 1.3 Repository has defined its designated community

2 The repository allows its designated community an adequate usage of the information represented by the digital objects.

- 2.1 Access for the designated community
- 2.2 Guarantees interpretability of the digital objects by the designated community

Example Criterion A 1.1

1.1 The digital repository has developed criteria for the selection of its digital objects. (→ Criterion)

The DR should have laid down which digital objects fall within its scope. This is often determined by the institution's overall task area, or stipulated by laws. The DR has developed collection guidelines, selection criteria, evaluation criteria or heritage generation criteria. The criteria may be content-based, formal or qualitative in nature. (→ General explanation of the criterion)

Example Criterion A 1.1

In the case of both state-owned and non-state-owned archives, the formal responsibility is generally derived from the relevant laws or the entity behind the archive (a state-owned archive accepts the documents of the state government, a corporate archive the documents of the company, a university archive, the documents of the university).

German National Library law - draft law approved by Bundesrat, Article 2 Tasks and authorisation: The Library is tasked with: 1. collecting, making an inventory of, analysing and bibliographically recording a) originals of all media works published since 1913 and b) originals of all foreign media works published in German since 1913, and ensuring the long-term preservation of these works, rendering them accessible to the general public, and providing central library and national library services.

Supported by the state libraries, the Baden-Württemberg online archive (BOA - http://www.boa-bw.de/) collects net publications …"which originate in Baden-Württemberg, or the content of which is related to the state, its towns and villages or inhabitants."

The Oxford Text Archive http://ota.ahds.ac.uk/ collects "high-quality scholarly electronic texts and linguistic corpora (and any related resources) of long-term interest and use across the range of humanities disciplines". The website contains a detailed "collections policy".

The document and publication server of the Humboldt University in Berlin collects "electronic academic documents published by employees of the Humboldt University" http://edoc.hu-berlin.de/e_info/leitlinien.php.
(→ Examples, comments, notes from different application areas, with no claim to exhaustiveness)

Example Criterion A 1.1

[Erpanet: Erpanet "Appraisal of Scientific Data" conference, 2003]

[Interpares Appraisal Task Force: Appraisal of Electronic Records: A Review of the Literature in English, 2006]

[Wiesenmüller, Heidrun et al.: Auswahlkriterien für das Sammeln von Netzpublikationen im Rahmen des elektronischen Pflichtexemplars: Empfehlungen der Arbeitsgemeinschaft der Regionalbibliotheken, 2004]

(→ Literature related to this criterion)

Further Work

- General
  - 2ed edition of the catalogue
  - Continuation of the work in a DIN WG
  - Conducting interviews / test audits
  - WS in January
- Standardization
  - Coaching, self-audit
  - Approach DIN / ISO
- Certification
  - Criteria must meet requirements of formal certification processes
  - Define an audit process
- Internationalization
  - Continuation of cooperation

The catalogue

- German Version
  - nestor–Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung: Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 1 (Entwurf zur öffentlichen Kommentierung), nestor Materialien 8, Juni 2006, Frankfurt am Main : nestor c/o Die Deutsche Bibliothek,
  - ***http://nbn-resolving.de/urn:nbn:de:0008-2006060710***

- English Version
  - nestor - Network of Expertise in Long-Term Storage of Digital Resources / Trusted Repository Certification Working Group: Criteria for Trusted Digital Long-Term Preservation Repositories, version 1 (Request for Public Comment),
  - ***http://nbn-resolving.de/urn:nbn:de:0008-2006060703***

- Information about the trusted repositories group at:
  - ***http://www.longtermpreservation.de/ag-repositories***

Thank you very much for your attention!
Comments? Questions?

Stefan Strathmann
Göttingen State and University Library
strathmann@sub.uni-goettingen.de
0049 - (0)551- 39-7806

Exercise

- Describe a relevant part of your organisations digital collection.
- Define the objects in this collection.

Exercise

- Describe the designated community of your archive.
  - Which access modalities would they expect?
  - What aspects of quality management do they expect?
  - What would they expect from your digital archive?
    - Authenticity related?
    - Integrity related?

Exercise

Conduced a test audit of your collection. Concentrate upon the first criterion. Keep in mind, the basic principles (documentation, transparency, adequacy, measurability) of the catalogue. To what measure are you able to fulfill the criterion? What is missing? Explain your ideas to fulfill the criterion.