14 March 2008, University of Tsukuba/Japan

筑波大学 University of Tsukuba

Japanese Symposium Digital Preservation

- Trusted Repositories and Preservation Policies -

Stefan Strathmann
Goettingen State and University Library
Germany
strathmann@sub.uni-goettingen.de

R&DD
Research & Development Department SUB Göttingen

Outline

- Preservation Policies
  - International Preservation Policy
  - National Preservation Policy
- Trusted Repositories
  - General introduction: Trusted Repositories
  - Initiatives and activities
  - Example: nestor Catalogue of Criteria for Trusted Digital Repositories

Preservation Policy

What are Preservation Policies?

- Digital Preservation is, on the one hand, a technical issue – on the other hand it is foremost an institutional, national and international organisational and political challenge!
- A Preservation Policy is the declared intention to preserve the digital heritage!
- A Preservation Policy is the plan to preserve!

Preservation Policy vs. Preservation Strategy

- A preservation strategy declares how digital preservation shall be reached (migration, emulation ...).
- A preservation policy declares which objects should be preserved, why, for whom and for how long.

The preservation policy is the basis for any preservation strategy!

Preservation Policy - Duration

- Preservation policies should be of long duration.
  - It should not be oriented on
    - technical innovation cycles
    - political changes
    - institutional changes

Preservation Policies - Examples

- International Preservation Policy
  - UNESCO Charta (2003)
  - Recommendation of the European Commission on the digitisation and online accessibility of cultural material and digital preservation (2006)
- National Preservation Policy
  - Australia
  - In Germany fragments and preparatory work exist (Law of the DNB, nestor Memorandum)
- Institutional Preservation Policy
  - National Archives of Canada
  - Online Computer Library Center – OCLC
  - National Archives (UK)

National Preservation Policy

- National framework for digital preservation
  - Not necessarily one single document, but a collection of laws, appointments, contracts, agreements etc.
- Dealing with several topics:
  - General commitment for the preservation of digital objects
  - Statement on availability and access
    - Digital preservation does not end in itself

National Preservation Policy

- Legal framework
  - Digital Preservation should be considered in legislation processes (e.g. copyright, archival law, personal rights etc.)
- Financial issues
  - A stable long-term financing must be established
- Responsibilities
- Selection Criteria

Development of a national preservation policy is a challenging and longsome process (broad consensus needed).

Institutional Preservation Policy

- Institutional area of application
  - Adjustment to the specific institutional needs
- Commitment
  - Internal: raising awareness
  - External: transparency => trustworthiness
- Day-to-day business must be adjusted to the policy (not vice versa; durability)
- Allocation of resources

Institutional Preservation Policy

- Precaution for the closing down of the institution (Fallback strategy)
- Using scenarios
  - What should be provided to whom in which way and with which regulations?
- Security
  - Often dealt with separately in IT-security documents

Institutional Preservation Policy

- Basis for the choice of the institutional preservation strategy (aligned to the collection development and the needs of the designated community)
- Sometimes very detailed policies with declaration of preservation strategies and technologies => often revised versions (OCLC)

Trusted Repositories

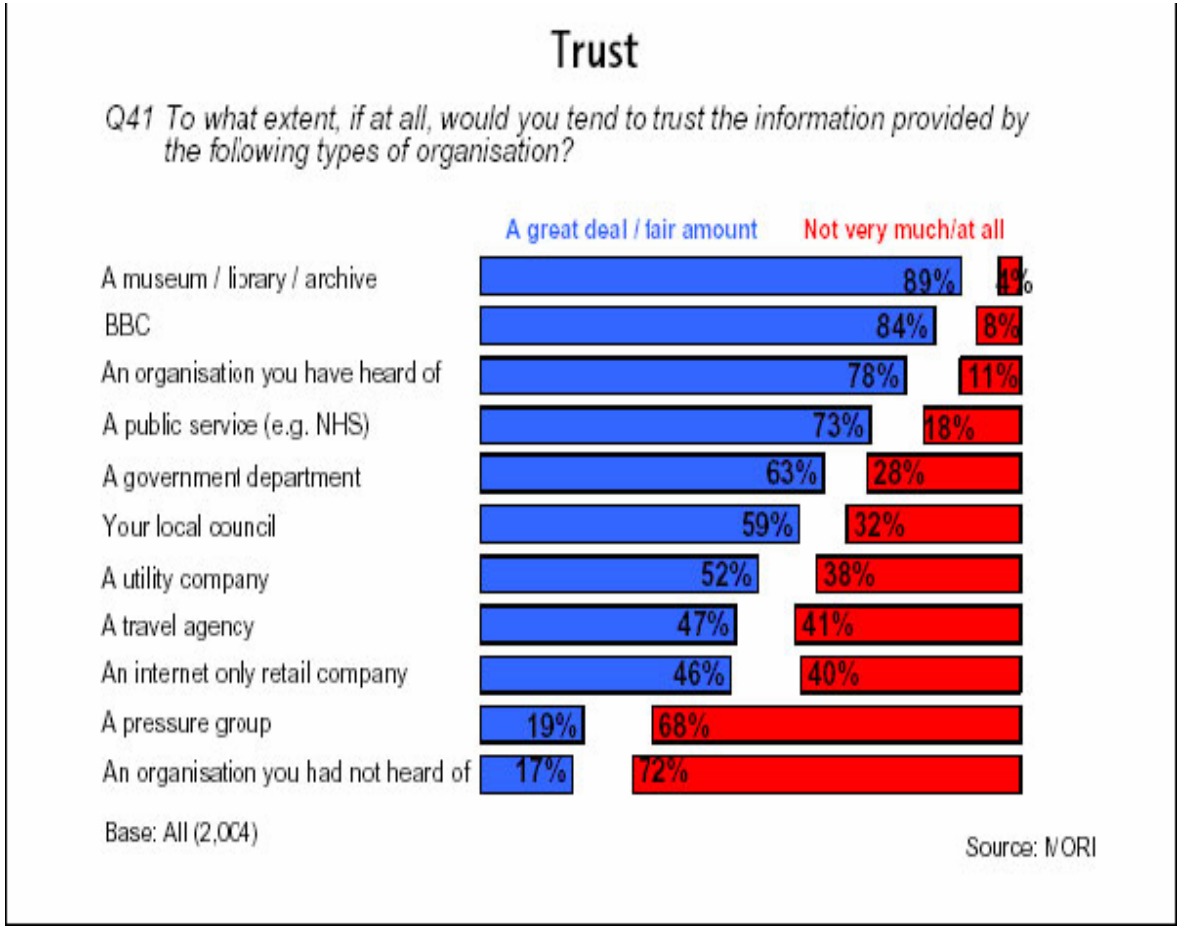
Digital Preservation & Trust

- Creation of digital information continues to accelerate!
- Digital information is 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



Source: Press Release: MORI survey uncovers major new trends in web use in the UK, 10 Feb 2005. See:

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)

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

- Revised and expanded version of “The Audit Checklist for the Certification of Trusted Digital Repositories”, originally developed by RLG-NARA
- 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

筑波大学 University of Tsukuba

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

DRAMBORA encourages repositories to:

- develop an organisational profile, describing and documenting mandate, objectives, activities and assets;
- identify and assess the risks that impede their activities and threaten their assets;
- manage the risks to mitigate the likelihood of their occurrence;
- establish effective contingencies to alleviate the effects of the risks that cannot be avoided.

(Andrew McHugh)

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 September 2006 – August 2009
- Funded by the German Ministry of Research and Education
- Partner: cultural heritage sector (libraries, archives, museums)
- Aim: information and communication - not archiving

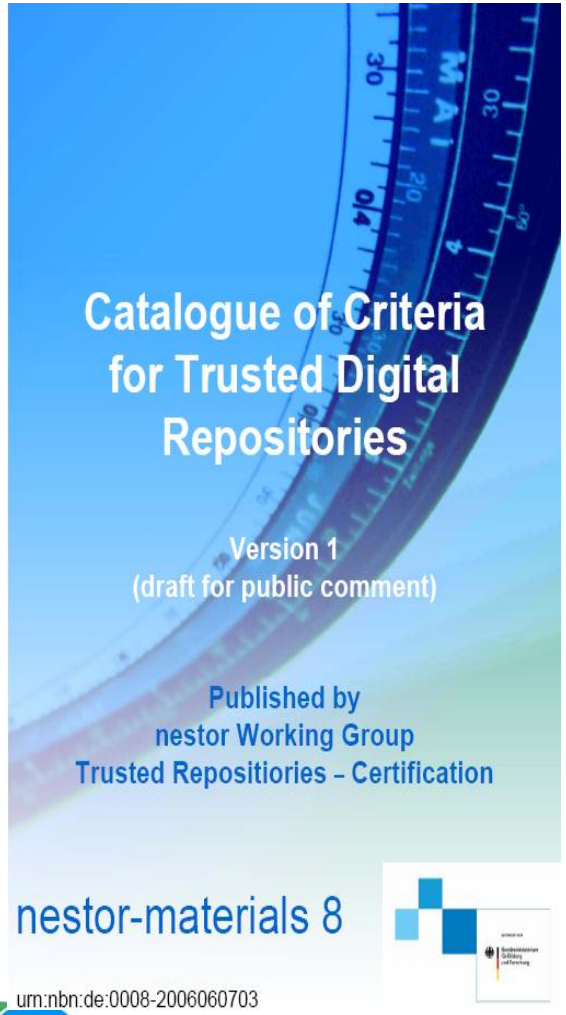
nestor WG on Trusted Repositories Certification

- 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

Catalogue of Criteria



- Public draft for comment in June 2006 (German version; English version in December 2006); at the moment: revision and enrichment
- Self-assessment tool
- Target group: cultural heritage organizations, software developers, third party vendors, …

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

筑波大学 University of Tsukuba

Catalogue of Criteria

- Formulates abstract criteria, enhanced with examples and explanations
- Focused on application in Germany, but orientated on international discussions and standards
- Catalogue of Criteria vs. Certification: After vital discussions, we assume it was too early for a formal certification process, and we wanted to start with the Catalogue of Criteria as a first step; it is intended to go on further (national/international standardization and formal certification)

Central Concepts of the Catalogue

- Key concept: Trustworthiness
  - A system operating according to its goals and specifications (it does exactly what it says)
  - From an IT security perspective: integrity, authenticity, confidentiality and availability

Central Concepts of the Catalogue

- 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 the Application

- Documentation
  - Allows to proof and evaluate the development of the system
- Transparency
  - Transparency to the outside
  - Transparency to the inside
- Adequacy
  - All criteria have to be seen in the actual preservation context
- Measurability
  - Partially no objectively measurable features
  - Indirect indicators can be made available (e.g. by transparency)

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)

Conclusion & Further Work

- Standardisation
  - Coaching, self-audit, testbed
  - Approach DIN / ISO
- Certification
  - Criteria must meet requirements of formal certification processes
  - Define an audit process
- Internationalisation
  - 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 nestor trusted repositories group at:
  - ***http://nestor.cms.hu-berlin.de/moinwiki/WG_Trusted_Repositories_-_Certification***

筑波大学 University of Tsukuba

Thank you very much for your attention!

Stefan Strathmann
Goettingen State and University Library
Germany
strathmann@sub.uni-goettingen.de

筑波大学
University of Tsukuba