The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification

Susanne Dobratz
Humboldt-University Berlin,
University Library
10099 Berlin, Germany
++49 30 2093 7070
dobratz@cms.hu-berlin.de

Dr. Astrid Schoger
Bavarian State Library, Digital Library
80328 München, Germany
++49 89 28638 2600
astrid.schoger@bsb-muenchen.de

Stefan Strathmann
Göttingen State and University Library
Papendiek 14
37073 Göttingen, Germany
++49 551 39 78 06
strathmann@sub.uni-goettingen.de

ABSTRACT

This paper describes the general approach *nestor – the German "Network of Expertise in Long-Term Storage of Digital Resources"* has taken in order to design a criteria catalogue for the self-assessment of trusted digital repositories used for long term preservation issues. Further developments will finally led to the implementation of a formal certification process for trusted digital repositories.

Categories and Subject Descriptors

J.7 [Computers in other Systems]

K.1 [The Computer Industry]

K.4.1 [Public Policy Issues]

K.6.1 [Project and People Management]

K.6.2 [Installation Management]

K.6.3 [Software Management]

K.6.4 [System Management]

K.6.5 [Security and Protection]

K.4.3 [Organizational Impacts]

K.4.4 [Electronic Commerce]

General Terms

Management, Measurement, Documentation, Reliability, Security, Human Factors, Standardization, Legal Aspects.

Keywords

Digital Libraries, Long Term Preservation, Certification, Trustworthiness, Digital Repositories

1. INTRODUCTION

One of the central challenges to long term preservation in a digital repository is the ability to guarantee the interpretability of digital objects for users across time. This includes a guarantee of integrity, authenticity, confidentiality and accessibility to the digital data. These attributes are compromised by the aging of storage media as well as rapid changes in the technical infrastructure. Malicious or erroneous human actions also put these digital objects at risk. An effective concept for trustworthy long-term preservation in digital repositories therefore considers both technical, as well as organizational, provisions and human resources. A trustworthy digital repository for long term preservation has to operate according to the repository's aims and specifications.

As the long-term preservation of digital objects is, globally speaking, in its infancy and little experience has been amassed to date, trustworthiness is not intended to, "… *give a declaration of guarantee for five or fifty years, but to enable institutions to develop strategies in order to cope with the continuous change of information technology in a responsible way"[1]*

2. Background

In December 2004, the German nestor project (Network of Expertise in Long-term STOrage of Digital Resources - A Digital Preservation Initiative for Germany) set up a working group on the certification of trusted digital repositories. The nestor working group consists of representatives from national, state and university libraries, federal and state archives, museums, data centers, publishers, and certification experts, from Germany and Austria. Taking into account the work of OCLC/RLG in 2002 *[4]*, the nestor group has focused on identifying features and ranges that may be relevant in evaluating digital object repositories (those that already exist as well as those which are just emerging or, as yet, are only planned). The aim is to form a web of trustworthiness in which those digital repositories can function as long-term digital archives within various environments: the library community, the archival world (in a traditional sense), the museum community, and other data producers such as government institutions, world data centers, and publishing houses.

In January 2005, the nestor group carried out a small-scale survey on recent standards and usage within digital repositories. It was followed by a public workshop in June 2005 and an expert round table in March 2006. A final report in the form of a criteria catalogue is due to be published in June 2006, see [2].

3. Defining the Target Group for the Criteria Catalogue

This criteria catalogue primarily addresses cultural heritage organizations, archives, libraries, and museums, and is designed as a guideline for the planning and setup of a digital long-term preservation repository. Secondarily, this catalogue can be an orientation guide for software developers, third party vendors, or service providers from the private sector.

Although the nestor catalogue is focused on application in Germany and it is crucial to analyze generally accepted criteria in regard to the situation in Germany, it must be discussed internationally and should adhere to international standards. In evaluating repositories, various components must be considered such as specific judicial constraints, the setup of public institutions (financially and in respect to human resources), national organizational decisions, and the status of development in Germany as a whole.

Potential interest groups for trustworthiness are:

- Repository users who want to access trustworthy information – today and in the future,
- Data producers, content providers for whom certification provides a means of quality assurance when choosing potential service providers,
- Resource allocators and funding agencies and institutions that need to make funding and granting decisions, and
- Digital long-term preservation repositories that want to or need to publicize their trustworthiness for the reason of competing for users and providers, and standing their ground against other competitors.

4. Coaching - Self-Audit – Certification

Currently, no method has been developed that would allow the formal certification of digital long-term preservation repositories according to this catalogue. This has been a vital point within the internal and public discussions of the *nestor Working Group on Trusted Digital Repository Certification.* We believe that it is still too early to introduce effective auditing.

For many of the abstract criteria expressed in the catalogue, it is not yet possible to define accepted standards on which auditing processes could be based. Therefore, nestor has for the moment focused on presenting the paper as a “guideline” for setting up a trusted digital repository. It is believed that this will be helpful for many institutions and will stimulate the development of trusted digital repositories. The catalogue can be used as an instrument for self-evaluation on all steps of the development from the concept via specification to implementation.

We regard that as the first step. Within a second step it is intended to participate in a national/international standardization process via DIN/ISO and to establish a formal certification process, where the catalogue will function as auditing tool.

Certification supports repositories that need to provide objective evidence and it encourages competition in the public sector. It supports the quality management and assurance of public administration.

Whenever raw data or research data has to be archived, a certification is considered to be very important.

5. Concepts Central to the Criteria Catalogue and the Evaluation of Trusted Repositories

5.1 Trustworthiness

The concept of trustworthiness is perceived as in “the common criteria in the evaluation and assessment of security in information technologies.” [12].

Digital objects are considered valuable assets that are endangered by decay or loss of integrity and authenticity.

Trustworthyness (German: Vertrauenswürdigkeit), see discussion in [8], is a feature that allows a system to operate according to it’s goals and specifications (it does exactly what it says).

From the IT security perspective integrity, authenticity, confidentiality and availability are important building blocks of trustworthy digital preservation repositories.

Integrity and authenticity support the completeness and elimination of unintentional modifications to repository objects, as well as protection of the objects’ integrity from malicious or erroneous human behaviour and from technical imperfection, damage, or loss of technical infrastructure. The principle of adequacy also applies to these preservation rules. Securing authenticity requires that the creator and the time of creation are identifiable and are securely received and stored. In the context of digital objects, this means that at the time of ingest the digital repository checks and verifies the identity of the depositor.

Availability or usability means that access to the repository by potential users is guaranteed. Another meaning is the guarantee that the objects within the repository are interpretable. The availability of the objects is defined as a central task that has to be fulfilled in relation to the designated audience and its requirements.

Under confidentiality shall be understood that the information objects can only be accessed by the permitted users.

The range of existing preservation repositories and those in development is wide, starting from national and state libraries and archives with deposit laws, via media centres having to preserve e-learning applications and hosting publication archives for smaller institutions to world data centres in charge of “raw” content data. For more examples see [2].

5.2 Steps towards a Trusted Digital Repository

A digital long term preservation repository occurs as a complex common coherence. The realization of single criteria has to be considered against the background of the overall system goals. The implementation of the digital long term preservation repository as well as the implementation of single criteria is executed as multi step process, containing the following steps:

1. Conception

2. Planning and Specification

3. Realization and Implementation

4. Evaluation

As preservation is regarded as a process underlying permanent changes, these steps cannot be taken as fixed model. In contrary they are repeated during the development when necessary. The development itself is controlled and monitored by a quality management.

Quality management [10] defines the quality goals of the digital long-term preservation repository. This includes a list of aims and responsibilities that allows for the definition and monitoring of an appropriate process structure. The quality management component defines all processes and their interdependencies, and verifies that responsibilities are assigned. This also applies to organizational external processes. Quality management provides an adequate procedure for documentation. The digital long-term preservation repository defines rules for completeness, correctness, actuality, understandability, and availability of the documentation, and implements those rules and controls for adherence. The quality management component enables the digital long-term preservation repository to adequately respond to substantial changes.

6. Basic Principles for the Derivation of Criteria

6.1 Abstraction

The catalogue's overall aim is to introduce stable criteria for a wide spectrum of possible applications and to maintain it's validity based on long term considerations.

For this reason the catalogue acts on the assumption that criteria have to be formulated on a very abstract level, to remain valid over time. They are enriched by detailed explanations and examples. The latter are conform to the current state of technology and organisation and may be reasonable within the context of a very special preservation task.

6.2 Accordance to OAIS Terminology

As basis for a common terminology and fort he structure of the catalogue the OIAS reference model was taken, where possible.

On one hand OAIS is used to describe core processes starting with ingest via archival storage to access. On the other hand OAIS allows to describe the life cycle of digital objects within the repository.

Following information packages have been considered: the Submission Information Package (SIP) for ingest, Archival Information Package (AIP) for the archival storage, and the Dissemination Information Package (DIP) for the usage.

The term digital object is regarded as defined in the OIAS information model.

7. Basic Principles for the Application of Criteria

7.1 Documentation

The goals, the concept, the specification as well as the implementation of a digital long term preservation repository must be documented adequately. The documentation demonstrates the development status internally and externally. A premature evaluation based on early documentation may also prevent mistakes and inappropriate implementations. An adequate documentation allows to prove the conclusiveness of the design and architecture of the digital long term preservation repository at all steps. In addition, all quality and security standards the require an adequate documentation, [11] .

7.2 Transparency

The transparency principle stands for the idea that functionalities of the digital repository are perceptible to the outside. Transparency in the outside levels the differences between insiders (e.g., reviewers, archivists) and outsiders (e.g., users, producers, data providers) and allows discussions between them. It supports the effort for trustworthiness and transparency so that:

Users can determine the level of trustworthiness themselves,

Creators are able to see where their objects are stored,

Repository funders can see what their money is spent for, and

Other digital long term preservation repositories, through this information, can enter into discussions and possible collaborations.

Transparency to the inside urges the need for documentation and enforces the necessity to actually fulfill proposed and published standards. It documents the operators, the management, the staff the adequate quality and ensured the traceability of taken measures. Transparency allow to restrict access to security relevant information.

Summarising: Transparency establishes trust, because it allows a direct evaluation of the quality of the digital long term preservation repository by different interest groups.

7.3 Adequacy

The adequacy principle includes the fact that no absolute evaluation of measurements is possible, but rather an evaluation has to consider the aims of the respective digital long term preservation repository.

The criteria have always to be seen within the actual preservation context, including that single criteria may become irrelevant under the specified goals and tasks.

7.4 Measurability

There are partially no objectively measurable features for trustworthiness, esp. under consideration of long term preservation issues.

In those cases on has to rely on indirect measures in order to evaluate the level of trustworthiness. So e.g. transparency can function as one instrument to make indirect indicators available for evaluation.

8. A Metric for Certification Criteria

Three examples of different approaches currently in use in Germany are presented. The DINI criteria distinguishes between minimum requirements and recommendations [6]. The DOMEA concept [7], used in the archives domain, works with requirement groups – basic requirements and specific requirements. Each can be rated in a range from 0 to 4 points. Within each group, a minimum amount of points must be achieved. The IT Grundschutzhandbuch (IT Basic Protection Manual), [9] published by the Federal Office for Information Security, uses an implementation status for each measurement.

Through several discussions, the nestor group came to the conclusion that a weighting of the different criteria should be avoided, since this is already implicitly included in the principle of adequacy. One could demand that all criteria of the nestor catalogue be fulfilled up to a certain level. Criteria that allow exceptions have to be marked and justified explicitly, whereupon the equality of alternatives has to be proven.

"Adequacy" as a metric for the evaluation of the fulfillment of a criterion is feasible, because "adequate" itself is a scale value that expresses, for a specific criterion, the optimum, independent from the archiving context. In the idea of the basic principle of efficiency, it means that neither too little nor too much has been done.

9. The Catalogue

Based on the initial nestor survey and similar to the RLG/NARA approach *[5]*, the group used abstract criteria in the main catalogue instead of asking very detailed and specific questions (e.g. which metadata is used). The nestor catalogue includes best practice values and provides examples and specific literature references for the listed criteria, despite the need to update such examples regularly. The intention is that this criteria catalogue, and its planned revisions, will help customers to share the same information and expectations.

9.1 Overview of Criteria

Within the following table the term "repository" is taken as abbreviation for "digital long term preservation repository" .

A	**Organisational Framework**
1	**The repository has defined it's goals.** 1.1 selection criteria 1.2 repository takes responsibility for the permanent preservation of the information represented by the digital objects 1.3 repository has defined target group / designated community
2	**The repository allows it's 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 digital object for designated community
3	**Legal and contractual rules are observed** 3.1 existence of legal contracts between producers and repository 3.2 repository operates on a legal basis regarding archiving 3.3 repository operate son a legal basis regarding usage
4	**The organisation chosen for the repository is adequate** 4.1 Adequate financing 4.2 Adequate staff qualification 4.3 Adequate organisational structure 4.4 repository has long term (strategic) plan 4.5 The continuation of preservation tasks is guaranteed even after existence of the repository
5	**An adequate quality management is conducted** 5.1 all processes and responsibilities have been defined 5.2 the repository documents all it's elements and processes 5.3 repository reacts against substantial changes

B	**Object Management**
6	**Repository ensures integrity of digital objects for all steps of processing** 6.1 Ingest 6.2 Archival storage 6.3 Access
7	**Repository ensures authenticity of digital objects for all steps of processing** 6.1 Ingest 6.2 Archival storage 6.3 Access
8	**Repository has a strategic plan for it's technical preservation strategies (preservation planning)**
9	**Repository transfers digital objects from it's producers following defined guidelines** 9.1 Repository specifies SIPs 9.2 Repository identifies relevant features of the digital objects for the information preservation 9.3 Repository has technical control over it's digital objects in order to execute preservation measures
10	**The archival storage of the digital objects is executed after well defined guidelines** 10.1 Repository defines it's AIPs 10.2 Repository ensures the transformation of the SIPS into AIPs 10.3 Repository guarantees storage and readability of the AIPs 10.4 Repository implements preservation strategies for AIPs
11	**Repository enables usage after well defined guidelines** 11.1 Repository defines it's DIPs 11.2 Repository ensures transformation of AIPs into DIPs
12	**The data management is suitable to guarantee the necessary functionality of the repository.** 12.1. Repository identifies it's objects and their relations uniquely and permanently

	12.2. Repository aquires adequate metadata for content and formal description and identification of the digital objects 12.3 Repository aquires adequate metadata for structural description of the digital objects 12.4 Repository aquires adequate metadata for documenting changes made on the digital objects 12.5 Repository aquires adequate metadata for the technical description of the digital objects 12.6 Repository aquires adequate metadata for the usage rights and terms of the digital objects 12.7. The assignment of metadata to the digital objects is guaranteed everytime

C	**Infrastructure and Security**
13	**The IT infrastructure is adequate** 13.1 The IT infrastructure implements the demands from the object management 13.2 The IT infrastructure implements the security demands of the object management
14	**The infrastructure ensures the protections of the repository and its digital objects**

9.2 Example Criteria

A criterion consists of 4 parts: the criterion itself, an explanation, possible examples and citations.

The security of the infrastructure is an example.

14 The infrastructure ensures the protection of digital long-term archives and its digital objects.

Explanation

The infrastructure protects the digital objects against system-dependent and external dangers. System-dependent dangers, for example, may be hardware problems or the loss of dedicated individual storage media (e.g., redundant storage). External dangers, for example, may be natural threats (e.g., fire, water, earthquake...). External dangers may also be the result of human error. Digital objects may be endangered, for example, by application viruses. To avoid damage, virus protection should be applied, whether risk results from a program (Trojaner) or from human intervention (espionage). Reference should be made to organization-sponsored basic protection manuals for ensuring digital object protection. Further, technical measures (e.g. virus protection programs) and organizational measures (e.g. admission regulations) should be implemented.

Examples

A backup, capable of taking over the enterprise with a measure of success, should be located at a remote location, to protect against natural or human disaster, such as a fire at a core building housing the digital repository hardware.

Appropriate technical safety precautions should be enforced. This will reduce access to protected data (e.g. archived documents of the state security service committee) to entitled users.

Access rights to digital objects are assigned and alloocated to specific persons/roles in the enterprise system level.

Literature

IT-Grundschutzhandbuch, http://www.bsi.de/gshb/

10. ACKNOWLEDGMENTS

We greatly acknowledge the whole nestor working group on trusted repository certification, see http://www.longtermpreservation.de/ag-repositories, esp. our collegues: Dr. Andrea Hänger, Karsten Huth, Max Kaiser, Dr. Christian Keitel, Dr. Jens Klump, Dr. Nikola Korb, Peter Rödig, Dr. Stefan Rohde-Enslin, Kathrin Schröder, Heidrun Wiesenmüller.

11. REFERENCES

[1] Ute Schwens, Hans Liegmann; Die digitale Welt – eine ständige Herausforderung: Rainer Kuhlen, Thomas Seeger und Dietmar Strauch (Eds) In: Grundlagen der praktischen Information und Dokumentation: Volume 1. Preprint, 5th edition. München, Saur, 2004. URL:http://www.langzeitarchivierung.de/downloads/digitalewelt.pdf

[2] 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), nestor Materialien 8, June 2006, Frankfurt am Main : nestor c/o Die Deutsche Bibliothek, urn:nbn:de:0008-2006060703, http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf (in press)

[3] Reference Model for an Open Archival Information System (OAIS) (2002): CCSDS 650.0-B-1: Blue Book – Consultative Committee for Space Data Systems, http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf

[4] RLG/OCLC Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report / Mountain View, CA : RLG, 2002. URL: http://www.rlg.org/en/pdfs/repositories.pdf (last viewed 28.04.2005)

[5] RLG-NARA Task Force on Digital Repository Certification: Audit Checklist for Certifying Digital Repositories. URL: http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf (last viewed 23.09.2005)

[6] Deutsche Initiative für Netzwerkinformation (*DINI*) Working Group Electronic Publishing: *DINI*-Certificate Document and Publication Repositories, November 2003, *DINI*-Schriften 3-en, urn:nbn:de:kobv:11-10046073.

[7] DOMEA-Koncept: http://www.kbst.bund.de/cln_006/nn_836960/Content/Standards/Domea__Konzept/domea__node.html__nnn=true

[8] Henry Gladney: Perspectives on Trustworthy Information, In Digital Document Quarterly (DDQ), Volume 1, Number 2, 1Q2002. URL: http://home.pacbell.net/hgladney/ddq_1_2.htm

[9] Bundesamt für Sicherheit in der Informationstechnik: Leitfaden IT-Sicherheit IT-Grundschutz kompakt, 2004, http://www.bsi.de/gshb (last viewed 28.04.2005)

[10] Quality Management DIN EN ISO 9000ff, Beuth-Verlag, 2006, CD-ROM

[11] ISO 15489-1 Information an d documentation – Records Management, 2001-09-15

[12] Bundesamt für Sicherheit in der Informationstechnik Common Criteria for Information Technologie Security Evaluation, Version 2.1, http://www.bsi.bund.de/cc/ccengl/downcc21.htm