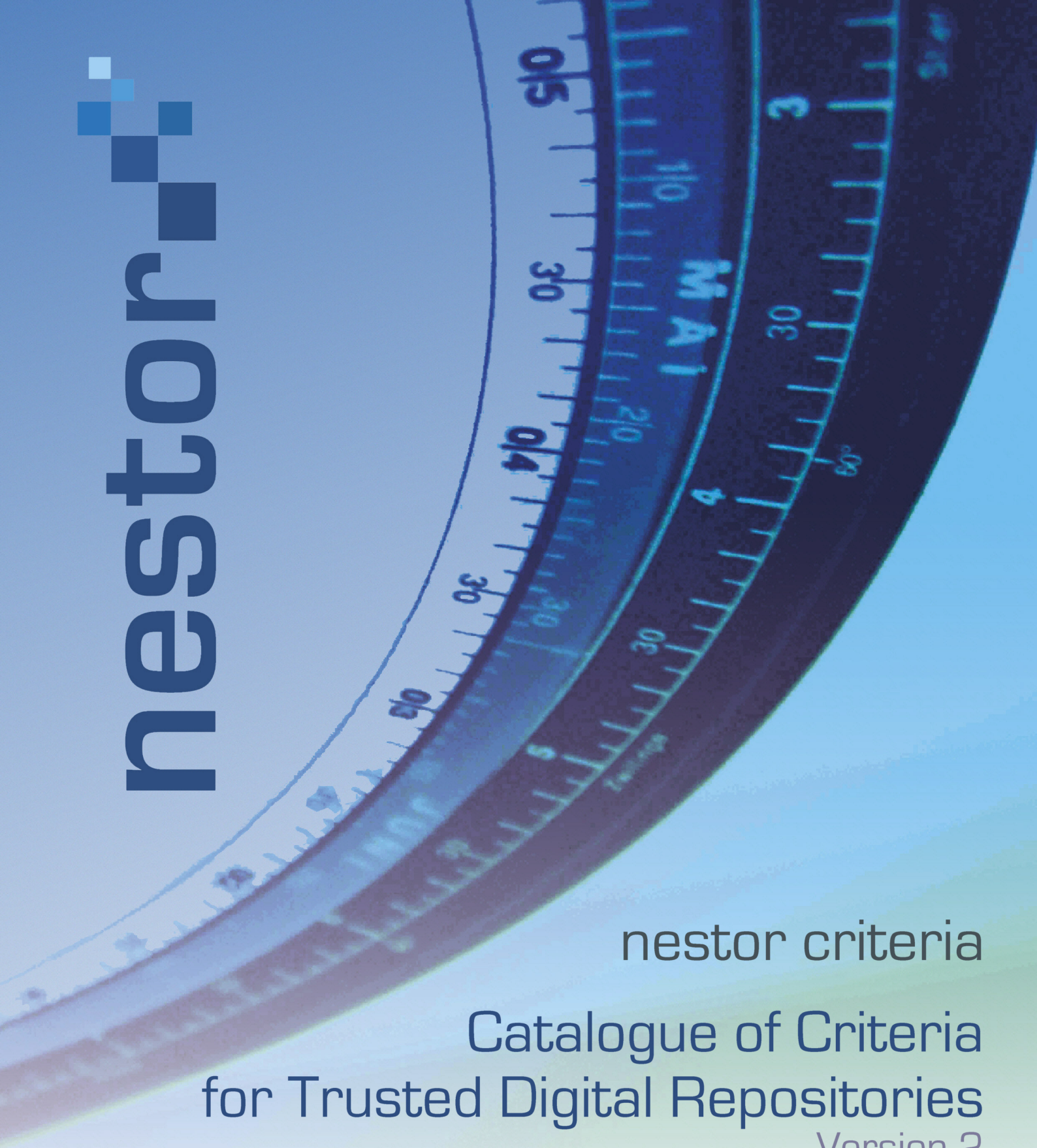

nestor

nestor criteria

Catalogue of Criteria for Trusted Digital Repositories

Version 2

published by nestor Working Group
Trusted Repositories - Certification

nestor-materials 8

nestor criteria

Catalogue of Criteria for Trusted Digital Repositories

- Version 2 -

published by
nestor Working Group
Trusted Repositories -
Certification

Frankfurt am Main, November 2009

nestor materials 8

Publishing details

nestor materials 8: nestor - Network of Expertise in long-term Storage and Accessibility of Digital Resources in Germany / Working group "Trusted Repositories – Certification": nestor criteria : Catalogue of Criteria for Trusted Digital Repositories, Version 2, 2009, Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek,
urn:nbn:de:0008-2010030806
http://nbn-resolving.de/urn:nbn:de:0008-2010030806

nestor working group „Trusted Repositories - Certification"

c/o Bayerische Staatsbibliothek, Digitale Bibliothek
Dr. Astrid Schoger
80328 München
Germany
Tel.: +49-89-28638-2600
Fax.: +49-89-28638-2672
E-Mail: astrid.schoger@bsb-muenchen.de

or

c/o Humboldt-Universität zu Berlin, Universitätsbibliothek
Susanne Dobratz
Unter den Linden 6
D-10099 Berlin
Germany
Tel.: +49-30-2093-7070
Fax.: +49-30-2093-2959
E-Mail: dobratz@cms.hu-berlin.de

nestor - Network of Expertise in long-term STORage and accessibility of digital resources in Germany

c/o Deutsche Nationalbibliothek
Reinhard Altenhöner
Adickesallee 1
D-60322 Frankfurt am Main
Germany
E-Mail: lza-info@langzeitarchivierung.de
Web: http://www.langzeitarchivierung.de

Authors of the Criteria Catalogue:

Bergmeyer, Winfried: Institut für Museumsforschung Berlin
Dobratz, Susanne: Humboldt-Universität zu Berlin, Universitätsbibliothek
Dr. Hänger, Andrea: Bundesarchiv Koblenz
Huth, Karsten: Bundesarchiv Koblenz
Dr. Keitel, Christian: Landesarchiv Baden-Württemberg
Dr. Klump, Jens: GeoForschungszentrum Potsdam
Dr. Korb, Nikola: Deutsche Nationalbibliothek Frankfurt
Rödig, Peter: Institut für Softwaretechnologie, Universität der Bundeswehr München
Dr. Rohde-Enslin, Stefan: Institut für Museumsforschung Berlin
Dr. Schoger, Astrid: Bayerische Staatsbibliothek München
Schröder, Kathrin: Bundesarchiv Koblenz
Steinke, Tobias: Deutsche Nationalbibliothek Frankfurt
Strathmann, Stefan: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Prof. Wiesenmüller, Heidrun: Hochschule der Medien Stuttgart

We are also grateful to the following for their suggestions:

Dr. Beckschulte, Klaus: Börsenverein des deutschen Buchhandels, Landesverband Bayern
Kaiser, Max: Österreichische Nationalbibliothek Wien
Dr. Lupprian, Karl-Ernst: Generaldirektion der Staatlichen Archive Bayerns
Dr. Schomburg, Silke: Hochschulbibliothekszentrum Köln
Ullrich, Dagmar: Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen

Input from the international discussion came from:

Dale, Robin: OCLC/RLG / University of California Santa Cruz (USA)
McHugh, Andrew: Digital Curation Centre (UK)
Reilly, Bernard: Center for Research Libraries (USA)
Prof. Ross, Seamus: Digital Curation Centre (UK)
Ruusalepp, Raivo: Digital Preservation Europe
Waltz, Marie: Center for Research Libraries (USA)

Contents

SUMMARY 1
I. Introduction 3
Long-term preservation of digital objects – Basic concepts 3
The nestor criteria catalogue 6
Basic principles for deriving criteria 6
Basic principles for applying the criteria 7
The nestor working group "Trusted repositories – Certification" 8
Current international approaches to evaluating trusted repositories 9
II. Criteria catalogue 10
A. Organisational framework 10
B. Object management 19
C. Infrastructure and Security 36
III. Checklist 38
IV. Glossary and abbreviations 41
V. Bibliography 44
Appendix 53
Core Requirements for Digital Archives [9] 53

SUMMARY

Digital information has become an indispensable part of our cultural and scientific heritage. Scientific findings, historical documents and cultural achievements are to a rapidly increasing extent being presented in electronic form - in many cases exclusively so. However, besides the invaluable advantages offered by this form, it also carries a serious disadvantage: users need to invest a great deal of technical effort in accessing the information. Also, the underlying technology is still undergoing further development at an exceptionally fast pace. The rapid obsolescence of the technology required to read the information combined with the frequently imperceptible physical decay of the media themselves represents a serious threat to preservation of the information content - not only in the long term!

These circumstances have given rise to justified questions regarding the trustworthiness of the information. Producers and consumers of information ask themselves which "memory organisations" are capable of ensuring the authenticity, integrity, confidentiality and availability of digital information. And, confronted with the inexorable flood of digital objects, those responsible in the institutions are faced with the challenge of earning trust and communicating this, be it to fulfil a legal mandate or simply to survive in the market.

This is the main focus of the work of the nestor working group "Trusted repositories - Certification". It identifies criteria which permit the trustworthiness of a digital repository to be evaluated, both at the organisational and technical levels. The criteria are defined in close collaboration with a wide range of different memory organisations, producers of information, experts and other interested parties. This open approach ensures a high degree of universal validity, suitability for daily practical use and also broad-based acceptance of the results. The present criteria catalogue represents an important milestone on the road towards achieving the working group's goals. The memory organisations should be given a well-constructed, coordinated and practical tool for achieving and proving their trustworthiness. However, the intention is also to present the option of documenting trustworthiness by means of certification in a standardised national or international process. The catalogue also supports active participation in existing international standardisation efforts.

In developing a broadly accepted criteria catalogue, nestor has used input and comment from the institutions which are affected by the issue or have a vested interest in it. In order to tackle the problems and to involve all interest groups from an early stage, the working group adopted an open procedure right from the outset. The current version also takes into account the ongoing discussions of our international partners, especially the OCLC/RLG-NARA Digital Repository Certification Task Force, the Center for Research Libraries, the Digital Curation Centre and the Digital Preservation Europe Project.

The present document first offers a brief introduction to the problems surrounding the long-term preservation of digital objects. A description of the main concepts and principles underpinning the criteria catalogue ensures comprehensibility. The aims and methods of the working group are briefly outlined. This introduction is followed by the criteria catalogue itself in its full, unabridged form. The present document is completed by a compact overview of the catalogue in the form of a checklist, plus a glossary.

I. Introduction

Long-term preservation of digital objects - Basic concepts

Threats to the preservation of information, trustworthiness

Information in the form of digital objects is threatened by impaired integrity, authenticity and confidentiality up to and including total loss of accessibility and usability. Hidden bonds between the information and the data carriers, physical ageing of the data carriers, and rapid changes in the technical infrastructure required to interpret the digital objects represent special challenges for long-term preservation. The purpose of digital repositories is to preserve information over long periods of time. They must therefore take both organisational and technical steps to counter these threats. Each trusted digital repository has its own targets and specifications. The trustworthiness of digital repositories can be tested and assessed on the basis of a criteria catalogue.

Information, representation, interpretation

An **information object** is a logically discrete unit of information. An information object can be represented (either in part or in full) by digital objects. This can be effected in a number of ways, i.e. by different representations (e.g. a literary work can be represented by a full-text digital publication as a text file, by a talking book as an audio file or by a retrospectively digitised version of a printed edition).
For the information to be used, the digital objects must be interpreted, allowing them to be processed by machines or understood by humans. This is carried out using so-called representation information (e.g. description of the text, image, audio format used, description of a database structure etc.). In order to maintain the information represented by digital objects, a digital repository must therefore preserve the representations together with the corresponding representation information.

The term information here covers all types of communicable knowledge content, e.g. works of intellectual creativity, results of research and development, documentation of political, social and economic events.

Digital objects are frequently organised into files. A digital object can be a single file (e.g. a digital photo saved as a TIFF file) or consist of a number of different files (e.g. an electronic journal consisting of individual articles saved as PDF files). In addition, a file may incorporate a number of digital objects (e.g. a database file).

Content data, metadata, digital objects

Data represents a formalised type of information which permits it to be interpreted, processed and exchanged. Digital objects are administration units of digital data in an IT system.
Besides data required for interpretation (representation metadata), further data may be added to the digital data representing an information object (content data) which helps e.g. identify, search for, document the integrity and authenticity of the content, and manage its usage rights. This so-called metadata can be created at different times in the lifecycle of digital objects (during production, archiving or provision for use etc.).
Together with the content data, it constitutes a conceptual unit – an information package. The information packages are subdivided into submission information packages (SIP), archival information packages (AIP) and disseminition information packages (DIP), depending on the processing stage in the digital repository.
During implementation, metadata can be stored and managed either separately (e.g. in a database system) or together with the content data (e.g. in a single file).

Digital repository

For the purposes of this criteria catalogue, a digital repository is an organisation (consisting of people and technical systems) which has assumed responsibility for the long-term preservation and long-term accessibility of digital objects, and also for their interpretability, for the purpose of their being used by a specific designated community. "Long-term" here means lasting beyond technological changes (to hard and software) and also any changes to the designated community. This definition of a digital repository is based on the reference model for an "Open Archival Information System" (OAIS) [1].[1]

The digital repository can be an element within a larger institution which also archives conventional objects. The connections between analogue and digital objects should be maintained and represented accordingly in the search.

Use by designated community

The preservation of information for future use is defined as a key task of long-term preservation which should be based on the designated community and its needs. Future use is contingent upon the integrity, authenticity, confidentiality and availability of the digital objects being preserved, but also on the designated community's ability to interpret the digital objects, thereby allowing them to reconstruct the information

contained in them in an appropriate manner. Legal or organisational changes and technical developments can lead to changes in the designated communities and their needs. A digital repository monitors these changes and reacts accordingly.

Trustworthiness

Trustworthiness is the capacity of a system to operate in accordance with its objectives and specifications (i.e. it does exactly what it claims to do). From the IT security perspective, the basic values are integrity, authenticity, confidentiality and availability. IT security is therefore an important prerequisite for trusted digital repositories.

The spectrum of existing digital repositories and those currently being set up is very broad, as shown by the following examples:

Example 1: A large universal library with responsibility for continually growing collections of digital publications from publishers and official sources, for scientifically relevant Internet resources and for the results of digitisation projects etc. The designated community of this digital repository is the general public. There are many different producers: publishing houses, digitisation centres, publishing institutions, private individuals etc. Such a library can also carry out long-term preservation services for smaller institutions; it may also be part of a network which permits cooperation with other libraries and grants users uniform access to cooperatively organised materials.

Example 2: A university library which, besides commercial scientific literature, also handles eLearning media, university publications, publications by university staff members etc. Here the users are students and university employees. The producers are mostly university staff members.

Example 3: A research institution which generates and archives large quantities of specialist data. Its designated community are scientists with the necessary specialist knowledge for interpreting this data.

Example 4: An archive which stores electronic documents from administrative organisations on the basis of legal archive requirements. Besides the general public, the main designated community are the producers themselves. Use may be prohibited over longer periods by means of protective rights.

Example 5: A museum which manages the digitisation of museum objects and also original digital art. Users are the general public, art experts, artists, etc.

Example 6: A service provider which carries out long-term preservation contract work for other institutions and their collections. The institutions themselves are responsible for building up the collections; the service provider offers reliable preservation of the digital objects, ensuring their availability and usability.

The road to creating a trusted digital repository

A long-term digital repository is a complex interrelated system. Implementation of the individual criteria must always be seen in the light of the objectives of the overall system. Both realisation of the long-term digital repository as a whole and also fulfilment of the individual criteria are multiple-stage processes:

1. Conception
2. Planning and Specification
3. Realisation and Implementation
4. Evaluation

These steps should not be regarded as a rigid phase model. Rather they must be repeated at regular intervals as the result of continuous improvements. Quality management is deployed to monitor this development process.

The nestor criteria catalogue

Users of the Criteria Catalogue

The present criteria catalogue is aimed principally at memory organisations (archives, libraries, museums) and serves as a manual for devising, planning and implementing a trusted digital long-term repository. It can also be used at all stages of development for self-checking.

Also, this catalogue is intended to provide orientation to all institutions which themselves operate an archive, commercial and non-commercial service providers, and third party providers of products.

Basic principles for deriving criteria

Abstraction

The aim of this catalogue is to formulate criteria which can be used for a broad spectrum of digital long-term repositories and which will retain their validity over a longer period. The assumption is therefore that relatively abstract criteria will need to

be chosen. The criteria are each accompanied by extensive explanations and concrete examples from different fields. The examples are state-of-the-art in terms of technology and organisation, although in some cases they may only make sense within the context of a particular archiving task. They make no claim to completeness.

Conformity with OAIS terminology

The OAIS reference model together with its functional entities and information model serves – where possible – as the basis for providing common terms and for structuring the criteria catalogue. The OAIS is used to describe the core processes from ingest of the digital objects into the digital repository, via archival storage through to access; on the other hand it is also used to describe the life cycle of digital objects from the producer via the digital long-term repository through to the user. For this the following information packages have been taken into account: submission information package – SIP for ingest, archival information package – AIP for archival storage and dissemination information package – DIP for access.

Basic principles for applying the criteria

Documentation

The objectives, the basic concept, specifications and implementation of the digital long-term repository should be documented. This permits a check of whether all the parts of the digital repository are fully coordinated. 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. All quality and security standards require a suitable documentation.

Transparency

Transparency is achieved by publishing appropriate parts of the documentation. External transparency for users and partners allows these to gauge the degree of trustworthiness for themselves. Transparency gives producers and suppliers the opportunity to assess for themselves to whom they wish to entrust their digital objects. Internal transparency documents to the operators, the backers, the management and also to the employees the appropriate quality of the digital repository and ensures that any measures taken can be traced.
Transparency can be restricted to a specified circle (e.g. certifying centre) for the parts of the documentation which are not suitable for the general public (e.g. company secrets, security-related information).

The principle of transparency raises trust levels as it permits interested parties to make a direct assessment of the quality of a digital repository.

Adequacy

The principle of adequacy derives from acknowledgement of the fact that no absolute standards are possible, rather that evaluation is always based on the objectives and tasks of the digital repository concerned. The criteria have to be seen within the context of each individual archiving task. Individual criteria may therefore prove irrelevant. The required degree of fulfilment for a particular criterion may differ depending on the objectives and tasks of the digital repository.

Measurability

In some cases – especially regarding long-term aspects – there are no objectively measurable features. In such cases indicators such as the existence of suitable metadata point to the level of trustworthiness. Transparency also makes the indicators accessible for evaluation.

The nestor working group "Trusted repositories – Certification"

The working group "Trusted repositories – Certification" has been set up within the BMBF (Federal Ministry of Education and Research) sponsored nestor project in order to create a first catalogue of criteria for trustworthiness and to prepare for the certification of digital repositories in accordance with nationally and internationally coordinated procedures. The members of the working group are representatives from libraries, archives, museums, research institutions, publishing houses, software and certification experts.

The present catalogue was developed in an open process, and involved other experts and institutions with practical experience in setting up digital repositories. It is based upon the results of a survey and an open workshop. The first draft of the criteria catalogue was presented and discussed at an experts' round table. The catalogue was published as a draft for public comment in June 2006. The large amount of feedback received has been incorporated in Version 2.
Version 1 was also published in English in December 2006 as a means of launching the nestor criteria catalogue into the international discussion.

The nestor catalogue has mainly been compiled for application in Germany, however it is also being discussed and standardised within the international context. Here it is crucial to identify generally valid criteria amongst the specifically national conditions. These lie e.g. in the legal framework, the provision of public institutions with adequate financial and human resources, in national organisational structures and the status of national development in the field of digital long-term preservation.

Current international approaches to evaluating trusted repositories

The nestor criteria catalogue takes into consideration national and international approaches and findings such as the DINI Certificate for document and publication servers [2], the RLG-OCLC report "Trusted Digital Repositories: Attributes and Responsibilities" (May 2002) [3] and the "Trustworthy Repositories Audit & Certification: Criteria and Checklist“ (TRAC) published by the OCLC/RLG-NARA Digital Repository Certification Task Force in draft form in August 2005 and then in its final version in February 2007. [4].

The *DRAMBORA* [5] tool was developed by the Digital Curation Centre (DCC) [6] in collaboration with the Digital Preservation Europe (DPE) [7] project. Based on the existing criteria catalogues it facilitates the analysis of a digital repository, right from its objectives and derived tasks through to the existing risks and countermeasures deployed, thereby permitting the skills, strengths and weaknesses of the system to be identified.

The working group also cooperated with the OCLC/RLG-NARA Digital Repository Certification Task Force and the "Certification of Digital Archives Project" of the Centre for Research Libraries (CRL), and the "Long-Lived Digital Collections“ project of the CRL [8] and the Digital Curation Centre and the EU-Project Digital Preservation Europe.

The international collaboration of the above initiatives led in January 2007 to the formulation of 10 core requirements for trusted repositories [9] (see the appendix).

For reasons of brevity the term "digital repository" is abbreviated to "DR" in the catalogue below. The masculine pronoun is occasionally used, in which case both women and men are always signified.

The following overview shows the structure of the criteria catalogue:

Criterion
General explanations of the criterion
`Examples, comments, notes from different application areas. There is no claim to completeness.`
Literature related to this criterion

II. Criteria catalogue

A. Organisational framework

The digital repository acts within an organisational framework which is determined by the defined goals, the legal conditions and the staffing and financial resources available.

1 The digital repository has defined its goals.

The DR should have a clear conception of its objectives. It has determined which tasks it fulfils, and which principles it observes in doing so. This is crucial, as trustworthiness is not an absolute term, rather it depends on the goals of the particular DR. Following the principle of adequacy, evaluation of the individual criteria is always based on the specific goals. The DR ensures that its objectives are transparent so that others – especially users and producers – can gauge the trustworthiness for themselves. (The goals are often published in the form of a Policy.)

PANDORA: The purpose of the PANDORA Archive, http://pandora.nla.gov.au/overview.html

Oxford Digital Library: Background, Services, Principles and Guidelines, http://www.odl.ox.ac.uk/about.htm

Dokumenten- und Publikationsserver der Humboldt-Universität zu Berlin: Ziele und inhaltliche Kriterien, http://edoc.hu-berlin.de/e_info/leitlinien.php

National Archives: Custodial policy for digital records, http://www.nationalarchives.gov.uk/recordsmanagement/custody/

National Archives and Records Administration (NARA, USA): ERA Vision Statement, http://www.archives.gov/era/about/vision.html

Beagrie, Neil: Digital Preservation Policies Study, http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

Erpanet-Conference: "Policies for Digital Preservation", 2003. [10]

1.1 The digital repository has developed criteria for the selection of its digital objects.

The DR should have laid down which digital objects fall within its ambit. This is often determined by the institution's overall task area, or is 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.

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).

Law regarding the German National Library (DNBG, http://www.bundesrecht.juris.de/dnbg/BJNR133800006.html)

§2 Tasks and authorisation:

The Library is tasked with the following:
1. a) for all media published in Germany since 1913 and
b) for originals of all foreign media published abroad in German since 1913, for translations of all German media into other languages and foreign-language media publications about Germany, the Library is responsible for collecting, making an inventory of, cataloguing and bibliographically indexing 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 (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).

Erpanet-Conference: "Appraisal of Scientific Data", 2003. [11]
Wiesenmüller, Heidrun; Jendral, Lars et al.: Auswahlkriterien für das Sammeln von Netzpublikationen im Rahmen des elektronischen Pflichtexemplars: Empfehlungen der Arbeitsgemeinschaft der Regionalbibliotheken, 2004. [12]

1.2 The digital repository assumes responsibility for long-term preservation of the information represented by the digital objects.

The DR explicitly declares its responsibility for the long-term preservation of the digital objects ingested as described under 1.1. Long-term preservation here means permanent retention of the usability of the information represented by the digital objects (cf. the OAIS information model).

Formulation on the website of the Internet Archive (http://www.archive.org/about/about.php): "The Internet Archive is working to prevent the Internet (...) and other "born-digital" materials from disappearing into the past. Collaborating with institutions including the Library of Congress and the Smithsonian, we are working to preserve a record for generations to come."

Formulation on the website of the Oxford Digital Library (http://www.odl.ox.ac.uk/principles.htm): "Like traditional collection development, long-term sustainability and permanent availability are major goals for the Oxford Digital Library."

1.3 The digital repository has defined its designated community/communities.

The general definition of the framework for a DR involves defining the designated community/communities. This includes knowledge of the specific

requirements of the designated community/communities influencing the selection of the services to be provided.
If the designated community/communities or its/their requirements change over time, the DR should respond by adapting its services.

Possible designated communities include:

- Employees of an official body, a research institute etc.
- Scientists working in a particular discipline
- The general public

2 The digital repository grants its designated community/communities adequate access to the information represented by the digital objects.

The DR should regard its primary task as the current and future use of the information represented by the digital objects on the part of its designated community. Use of the digital objects is predicated on their preservation, their accessibility and on ensuring the ability to interpret them. Use may be adequate despite being restricted due to legal reasons (c.f. 3.3.) or not all properties of the original being preserved (c.f. 9.2.).

So-called "dark archives" are set up with no possibility for their use at present; they will only be used if the primary archive is rendered non-functional for whatever reason. In the event of such a crisis, use must then be guaranteed.

2.1 The digital repository ensures its designated community/ communities can access the digital objects.

The DR should ensure that authorised users have access to the digital objects. This includes appropriate search possibilities. When determining its service portfolio, the DR takes the needs of its designated community/communities into account. The DR announces in advance its conditions of use and any costs which may arise, listing these in a transparent manner.

Access can be obtained by:

- Accessing the digital objects
- Creating or supplying an analogue copy (e.g., as print-out by the user or in the form of a print-on-demand service)
- Creating or supplying a digital copy (e.g. download to a storage medium by the user, email delivery)
- Creating interfaces to permit access via other systems to the digital objects.

2.2 The digital repository ensures that the designated community/communities can interpret the digital objects.

The DR should take appropriate measures to ensure that the digital objects can be interpreted on a long-term basis, thereby creating the basic requisites for appropriate use. This includes the ability to interpret both content and metadata.
In ensuring this, the DR should take the needs of its designated community/communities into account. The more specialised the designated community/communities, the more know-how and technical equipment (e.g. certain software) is required, or the greater the willingness to organise additional equipment (installation of plug-ins) must be.
Changes to the technical environment or the designated community or communities can influence the ability to interpret the objects. The DR therefore should check at regular intervals, using appropriate procedures, whether the objects can still be interpreted by the designated community or communities.

- Possible measures include:
- Conversion into a current standard format
- Provision of emulation software (e.g. open source DOS emulator "DOSBox")
- Provision of representation information: e.g. documentation of data structures and field content to ensure that users can transfer data from specialist applications (databases) into the respective current database applications
- Provision of instructions for use, installation instructions, help texts
- DR carries out research or evaluation work as (charged) service
- Checking of interpretability on the basis of regular spot checks
- Provision of a feedback form which users can use to register interpretation problems

3 Legal and contractual rules are observed.

The DR's actions should be based on legal regulations. These cover the acquisition of the digital objects and also their archiving and use. Here the DR should strike a balance between the legitimate interests of the producers, those of the users and also, where applicable, the individuals concerned (in the case of person-related data).

BArchG: Gesetz über die Sicherung und Nutzung von Archivgut des Bundes, 2005. [13]
Goebel, Jürgen W.; Scheller, Jürgen et al.: nestor - materialien 1: Digitale Langzeitarchivierung und Recht, 2004. [14]
Coyle, Karen: Rights in the PREMIS Data Model. A Report for the Library of Congress, 2006. [15]

3.1 Legal contracts exist between producers and the digital repository.

In order to ensure planning and legal security the DR, where possible, should conclude formal agreements with the producers or suppliers. The nature and scope of the delivery is regulated, as are the DR's archival obligations, the conditions of use and, where applicable, the costs. The legal agreements should be supplemented with concrete implementation provisions. If it is not possible to conclude a formal agreement, the grounds for this should be given.

Possible agreements include:
a) Laws, regulations: Deposit Law, Archive laws
Law regarding the German National Library (DNBG)
http://www.bgblportal.de/BGBL/bgbl1f/bgbl106s1338.pdf
b) Contracts, agreements:
Licence agreements (cf. archiving clause in JISC model contract for electronic journals:
http://www.nesli2.ac.uk/NESLi2_licence_journals_final011003.htm)
Framework contracts
Deposit contracts, archiving agreements, archiving and use permits (cf. Bayerische Staatsbibliothek: http://www.babs-muenchen.de/content/netzpublikationen/einzelbewilligung.pdf)
Such an agreement, or its implementation clauses, define e.g.:
i) in which form the cooperation between producer / supplier and the DR should take; how communication is organised.
ii) the type and scope of delivery:
Scope, schedules, transfer procedure (data carrier, file transfer via networks, upload, download), file formats, other file properties (e.g. without active elements), additional information (e.g. the content and structure of descriptive metadata, XML schema etc).
iii) the DR's obligation:
Time of assumption of legal responsibility, duration of archiving, use of preservation measures (multiple copies, change-causing actions e.g. during migrations), significant characteristics.
iv) the conditions of use:
designated communities, services offered, usage rights, costs.

There is no possibility of a formal agreement regarding the archiving of STASI documents as neither a rights holder nor a legal successor exists.

3.2 In carrying out its archiving tasks, the digital repository acts on the basis of legal arrangements.

The DR should take legal provisions and contractual obligations into consideration regarding its archival storage and the use of preservation measures.

Restrictions imposed on archival storage by copyright can e.g. be countered by explicit agreements on the right to multiple storage, file-altering actions etc.

3.3 With regard to use, the digital repository acts on the basis of legal arrangements.

The DR should take legal provisions and contractual obligations into consideration regarding the use of digital objects. If this results in restricted use, the reason for the restriction should be documented.

Legal arrangements which can influence use include copyright, data protection, other legal regulations (e.g. periods of copyright for archives), contractual obligations or the contractual or legal stipulation of use.

Restrictions on use can be countered in some cases by controlled access to the digital objects. For the observance of copyright restrictions this could involve registration, exclusive use on the premises or on the Intranet or through charged usage/billing models. Separate declarations of commitment or the issue of anonymised user copies are possible options for compliance with data protection and archive regulations.

4 The organisational form is appropriate for the digital repository.

The DR should be organised in such a way that it can fulfil its short, medium and long-term goals. Its effectiveness and sustainability can be evaluated by users and producers. This evaluation is based on the following points.

Erpanet-Conference: "Business Models related to Digital Preservation", 2004. [16]

4.1 Adequate financing of the digital repository is secured.

The DR should be able to substantiate its claim that the proposed services can be financed, both in the short and long term.

The financing of the digital repository should have a legally secured basis.
In the case of state-financed digital repositories, the financing should be included in the formal planning documents (at least medium-term).
A private DR should be able to guarantee its financial sustainability on the basis of charged use of its services and on a long-term business plan.

Digital Longevity Department: Vers van de pers...Kostenmodel digitale bewaring, 2006. [17]
Palm, Jonas: The Digital Black Hole, 2006. [18]
Oltmans, Erik; Kol, Nanda: A Comparison Between Migration and Emulation in Terms of Costs, 2005. [19]

4.2 Sufficient numbers of appropriately qualified staff are available.

The qualifications and training of the staff should be adequate for the goals, tasks and processes of the DR. Suitable schemes should be in place to ensure adequate training and further training in the long term. Staff numbers

should be sufficient to allow all necessary processes to be completed in full. The long-term planning of the DR includes staffing resources.

Staff development includes task-based initial and further training, e.g. through courses and the provision of appropriate literature.
One aspect of training is active participation in relevant national and international conferences and working groups plus work on standardisation bodies. Such active participation makes the training levels of the staff externally visible.
Any shortfall in internal capacity can be compensated by external capacity.

4.3 Appropriate organisational structures exist for the digital repository.

The organisational structure should be adequate for the targets, tasks and processes of the DR. The processes and the allocation of staff and other resources are structured in such a way that the defined goals can be met.

The DR is incorporated at the appropriate point in the business distribution plan.
An organisational chart illustrates this relationship.

4.4 The digital repository engages in long-term planning.

The DR should engage in pre-emptive planning, including imminent or expected tasks, plus specify the deadlines by which they are to be completed. The management should have suitable structures and procedures for strategic planning. The basis for long-term planning is monitoring of legal and social changes, the demands and expectations of the target groups (in OAIS: "Monitor Designated Community") and all technical developments (in OAIS: "Monitor Technology") relevant for the sustained preservation and appropriate use of the information represented by the digital objects. The planning also includes securing the necessary resources.

Relevant legislative processes should be monitored right from the early phases (e.g. law on basic conditions for electronic signatures).
Strategic planning requires access to reliable current data. Process cost accounting, for instance, helps long-term planning of the required resources.

4.5 The digital repository reacts to substantial changes

Substantial alterations are those after which the goals can no longer be fulfilled unless reacted to, or their fulfilment at least carries an increased risk. Substantial alterations can either be technical, organisational or community-based.

For this the management should incorporate a process element which monitors changes, evaluates possible effects on task fulfilment and plans,

implements and monitors any necessary alterations.

The monitoring of technical developments includes e.g. the development and standardisation of new file formats and new storage techniques and, accordingly, any obsolescence of existing technologies arising as a result.
A substantial technical change could be a fundamental change in human-machine-communication.

A substantial change affecting the organisation as a whole could be the loss of a backer and therefore the financial base.

4.6 Continuation of the preservation tasks is ensured even beyond the existence of the digital repository.

The DR should have made contingency plans. In such a case the preservation work must be continued in a different organisational framework, thereby ensuring that the set tasks can be carried out in full. Where this is not possible, any restrictions should be documented. The DR should take precautions to ensure that the transition process can be defined, planned and implemented in good time. Suitable documentation is the basis for the success of a possible transition process.

This includes exportability of all the archive packages (including metadata) in a form which can be interpreted by the successor, as a means of guaranteeing the interpretability and authenticity of the data.
An external or higher body should guarantee continuation of the defined tasks.
Continuation should be governed by an agreement with a comparable organisation.

5 The digital repository undertakes appropriate quality management.

Quality management should ensure that the DR's goals are reached. The general targets should be broken down into specific aims and objectives. Suitable process structures should be established for this which are monitored by the quality management system.
The quality management should be a cross-sectional process covering all parts of the DR.

ISO 9000:2005: Quality management systems. Fundamentals and vocabulary, 2005. [20]
Liggesmeyer, Peter: Software-Qualität, 2002. [21]
Kneuper, Ralf: Verbesserung von Softwareprozessen mit Capability Maturity Model Integration, 2006. [22]
ITIL: IT - Infrastructure Library. [23]

5.1 All processes and responsibilities have been defined.

The quality management system should ensure that all processes and their interactions are defined, in particular that specified individuals are assigned responsibility for all processes. This also applies to external (outsourced) processes.

It is easier to determine the completeness of the processes and their interactions if a suitable reference model is available. The OAIS functional entities of Ingest, Archival Storage and Access can be used as the basis for defining the core processes. Support and management processes (data management, quality management, etc.) can then be defined on the basis of these core processes.

External processes require an internal process which contractually defines the services and a process which checks them. Responsibility should be assigned for these internal processes.

CCSDS: Producer-Archive Interface Methodology - Abstract Standard, Blue Book, 2004. [24]
Erpanet-Conference: "Workshop on Workflow", 2004. [25]

5.2 The digital repository documents all its elements based on a defined process.

The elements include: targets, plans, specifications, implementations, processes, software, objects and metadata etc.
The quality management system should include a suitable procedure for documentation, i.e. a system to manage all necessary documents. The DR should lay down rules regarding the completeness, correctness, validity, comprehensibility and accessibility of the documentation, implement these and monitor their observance.
This avoids knowledge being tied to certain individuals.

Standardised terminology which is adapted to the needs of the documentation users helps improve comprehensibility, for instance. Accordingly the documentation can be formal (e.g. for description of critical software processes), semi-formal (for conceptual description of processes and IT infrastructure) or natural (e.g. for external description of archive's objectives).

- Software documentation
- Process documentation
- Documentation of object formats

B. Object management

The digital repository should analyse its goals and strategies, and specify all object-related requirements for digital object management during the lifecycle of the objects in the DR. The main phases correspond in the OAIS reference model to the processes ("functional entities") of ingest, archival storage, including implementation of long-term preservation measures, and access. Additions to these functions may become necessary depending on the goals of the digital repository. Object management is based on the information model of the OAIS reference model and defines appropriate submission information packages (SIPs), archival information packages (AIPs) and dissemination information packages (DIPs). [26]

The integrity and authenticity of the information to be received are core concepts of trustworthiness. Integrity and authenticity should therefore be comprehensively ensured in all phases for which the DR has assumed responsibility. The prerequisite for this is ensuring the integrity and authenticity of the digital objects representing the information being preserved (cf. 6 and 7). The DR undertakes object-based planning of the long-term archiving measures to preserve the information (in OAIS: Preservation Planning) (cf. 8). Specified standards for ingest, archival storage and use, and standards for the objects themselves and their transformations are further indicators of trustworthiness (cf. 9, 10 and 11). The DR carries out transparent, metadata-based data management in order to reconstruct usable information from the digital objects for the designated communities and to fulfil requirements for integrity, authenticity and legally approved use of the information (cf. 12).

Object management requirements are the prerequisites for planning and operating the technical infrastructure and security system (cf. 13, 14).

6 The digital repository ensures the integrity of the digital objects during all processing stages.

Integrity here refers firstly to the completeness of the digital objects and secondly to their intactness.
The yardsticks for integrity are the characteristics of a digital object defined as worthy of preservation (cf. 9.2.).

Risks to integrity are posed by human activity (malicious or accidental), technical imperfection or theft of technical infrastructure.

The DR should take both organisational and technical precautions to ensure the integrity.
The DR should operate a data management system suitable for preserving integrity for the processes of ingest, archival storage and access. The DR should also take precautions regarding the integrity of the data management itself.

In exceptional cases the integrity may be compromised, in which case this must be adequately documented.

An example of deliberate or accidental modification is the input of virus-infected objects, the execution of which can result in changes or modifications to objects or other system elements (e.g. database scripts which delete objects or metadata).
Examples of technical imperfections are: faulty or incomplete software, especially that used for complex transformations (migrations), and storage media which is obsolete or has not been stored in conformity with the specifications. Generally, however, technical imperfections which are foreseeable should be remedied or flagged by means of appropriate error correction or error identification procedures. In some cases the user can select higher level error correction procedures for certain system components (e.g. through a higher degree of redundancy). This should be made full use of, where applicable.
The integrity of the objects may be impaired during the archiving of websites due to the functionalities of the offline browsers used (e.g. lack of moving images), however a protocol should document this.

ISO 15489-1:2001: Information and documentation – Records Management, 2001. [27]
Shirey, Robert W.: Internet Security Glossary, 2000. [28]
ISO/IEC 15408-x:2005: Information technology. Security techniques - Evaluation criteria for IT security, 2005. [29]
DigiCult: Integrity and Authenticity of Digital Cultural Heritage Objects, 2002. [30]

6.1 Ingest: the digital repository ensures the integrity of the digital objects.

For this the DR specifies a clearly identified interface to the producer and the archival storage. This includes the transformation from submission information packages into archival information packages. The interface allows the producer and the DR administration to check and maintain the integrity of the digital objects.

The DR should have drawn up agreements with e.g. its producers and/or suppliers regarding the technical aspects of the submission (ingest) transactions. In particular, there should be an agreement governing the transfer of responsibility for the integrity of the objects.
The DR should agree with the producer/supplier which properties digital objects need in order to exclude any risks to their integrity. Such measures could be e.g. prior removal of executable codes in digital documents.
The DR should ensure secure transfer channels from the producer or

supplier to the DR.
The DR should conduct checks to ensure the completeness and quality of the deliveries.

CCSDS: Producer-Archive Interface Methodology - Abstract Standard, Blue Book, 2004. [24]
Littman, Justin: A Technical Approach and Distributed Model for Validation of Digital Objects, 2006. [31]

6.2 Archival Storage: the digital repository ensures the integrity of the digital objects.

Here the DR should specify all the archival storage functions which are required to check and maintain the integrity of the digital objects by the administration of the DR. The functions include recording the archival information packages onto storage media, permanent storage, restoration of the archival information packages and all changes to the AIPs.

The DR should have a transparent procedure for determining the required degree of physical redundancy and suitable locations for storage media and related subsystems.
The DR should stipulate the required quality of the storage media (e.g. the use of standardised and certified storage media).
The DR should have laid down a policy regarding logical access to the archive store; this should include internal DR users, such as system administrators.
The DR should strictly regulate physical access to the IT systems.
The DR should have regulations concerning refreshment of the media or migration of the digital objects to other media.

6.3 Access: the digital repository ensures the integrity of the digital objects.

For this the DR specifies a clearly identified interface to the user and the archival store. This includes the transformation from archival packages into access packages. The interface allows the user and the administration of the DR to check and maintain the integrity of the digital objects.

The DR should ensure that no unauthorised user can obtain rights over digital objects, metadata or other system elements.
The DR should define how far its responsibility for the integrity of the digital objects extends in the delivery process.
The DR should analyse the quality of the interpretation aids it provides and make the results available to the users.

7 The digital repository ensures the authenticity of the digital objects during all stages of processing.

Authenticity here means that the object is genuine, i.e. that it represents, what it claims to represent. A key aspect is that the object in question was created by the given source and at the given time. Authenticity also includes full documentation of all transformations to the objects carried out for the purpose of preservation.

The DR should document if authenticity cannot be demonstrated in a particular object. Once the object has been received by the DR, the DR assumes responsibility for its authenticity.

The DR should operate a data management system suitable for preserving authenticity in the processes of ingest, archival storage and access. This is provided in particular by documentation of all changes to the objects (including metadata) (see 12.4).

```
Authenticity means that the producer or sender and the given production
or transmission time correspond to the facts, e.g. that an eMail
supposedly generated and transmitted by a particular person at a
particular time is actually from this person and was sent at the given
time.
```

ISO 15489-1:2001: Information and documentation – Records Management, 2001. [27]
Shirey, R.: Internet Security Glossary, 2000. [28]
ISO/IEC 15408-x:2005: Information technology. Security techniques - Evaluation criteria for IT security, 2005. [29]
PREMIS Working Group: Data Dictionary for Preservation Metadata, Version 2.0 2008. [32]
Gladney, H. M.; Bennett, J. L. : What Do We Mean by Authentic? What's the Real McCoy?, 2003. [33]
InterPARES. [34]

7.1 Ingest: the digital repository ensures the authenticity of the digital objects.

The DR should specify methods for assessing and securing the authenticity of the submission packages.

```
The DR should demand the formal registration of producers/suppliers with
an authorised body.
In certain contexts the use of digital signatures can ensure the
authenticity of the transfer packages.
The DR should require the producer/supplier to install procedures to
assess the authenticity of the digital objects, e.g. on the basis of
metadata describing the origin.
```

CCSDS: Producer-Archive Interface Methodology - Abstract Standard, Blue Book, 2004. [24]

7.2 Archival Storage: the digital repository ensures the authenticity of the digital objects.

The DR should specify methods which ensure the authenticity of the objects during implementation of the long-term preservation measures, or document the degree of authenticity (cf. 10.4 and 12.4).

The DR should keep full documentation of all transforming (i.e. altering or deleting) operations on the digital objects.

7.3 Access: the digital repository ensures the authenticity of the digital objects.

The DR should ensure the authenticity of the access packages and allow the user to determine the degree of their authenticity. In addition, the DR should also authenticate itself to the user as the supplier of the access packages.

The DR should provide the user with metadata which documents the origin and all changes in the archiving process, thereby permitting evaluation of authenticity.
The DR should register with an authorised body, e.g. the regulator for postal and telecommunications affairs, from which it receives a digital signature key certificate which it uses to generate digital signatures.
In certain contexts digital signatures can be used to ensure the authenticity of the access packages.
Compare also the method deployed in the ArchiSafe project by the Physikalisch-Technische Bundesanstalt Braunschweig for the use of digital signatures (http://www.archisafe.de/s/archisafe/index).

8 The digital repository has a strategic plan for its technical preservation measures (preservation planning).

In order to fulfil its responsibility for preserving information, the DR should have a strategic plan covering all outstanding or expected tasks, and the time of their realisation. This strategic planning (cf. 4.4) should be specified at the object level. Such measures should keep pace with ongoing technical developments (changes to data carriers, data formats, user demands etc.).

Measures for the physical preservation of the data (integrity, authenticity), its accessibility and the preservation of its interpretability should be used for the long-term preservation of the information represented by digital objects. Long-term preservation measures cover both content and metadata.

See 10.4 regarding implementation of the long-term preservation measures.

Output onto analogue media (e.g. microfilm) and redigitisation may be appropriate for certain digital objects.
The following are the main methods used to preserve interpretability:

Conversion to a current format or a current format version (migration, e.g. transformation of file in PDF 1.4 format to PDF/A format)
Recreation of the old application environment within a new technical infrastructure (emulation, e.g. DIOSCURI, http://dioscuri.sourceforge.net/
Long-term planning of the tasks arising from the formats can be based e.g. on a format register. Format registers are currently being developed by e.g. Harvard http://hul.harvard.edu/gdfr/ and the National Archives, Kew (PRONOM: http://www.nationalarchives.gov.uk/pronom/).
The decision-making process can be supported by the method (and in the near future the corresponding tool PLATO) developed as part of the PLANETS project.

nestor - Handbuch : Digitale Erhaltungsstrategien, 2008. [35]
Rauch, Carl; Rauber, Andreas: Anwendung der Nutzwertanalyse zur Bewertung von Strategien zur langfristigen Erhaltung digitale Objekte, 2006. [36]
Strodl, Stephan; Becker, Christoph et al.: How to Choose a Digital Preservation Strategy: Evaluating a Preservation Planning Procedure, 2007. [37]
Oltmans, Erik; Kol, Nanda: A Comparison Between Migration and Emulation in Terms of Costs, 2005. [19]

9 The digital repository accepts digital objects from the producers based on defined criteria.

The general collection guidelines, selection criteria, evaluation criteria or criteria for heritage generation (cf. 1.1) and the general aims of long-term preservation should be specified at the object level.
The transfer can be effected by submission of the objects to the DR by the issuing party or through manual or automatic collection on the part of the DR.

DOMEA : Aussonderung und Archivierung elektronischer Akten. Erweiterungsmodul zum Organisationskonzept 2.1, 2005. [38]
The U.S. National Archives & Records Administration: Disposition of Federal Records. Subpart L -- Transfer of Records to the National Archives of the United States, Part 1228, § 1228.270, 2002. [39]
NDAD: Transfer Procedures (Overview), 2005. [40]
DPC: Decision Tree for Selection of Digital Materials for Long-term Retention, 2006. [41]
nestor: Wege ins Archiv: Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv. [42]

9.1 The digital repository specifies its submission information packages (SIPs).

The DR should specify, or agree with the producers or suppliers, which digital objects and metadata are to be ingested into the DR (in the conceptual unit of a submission information package). These agreements should allow the transfer or the collection to be automated, and workflows for submission

to the DR to be implemented.
These specifications are the basis for quality checking the transfer objects.

Transfer packages may contain content data and also metadata, e.g. to establish their authenticity.
In the case of harvesting based on offline browsers, only text content - but not audio, video and other multimedia content - is collected (through the selection or exclusion of specific file formats).
The file formats of the transfer packages can be validated using JHOVE (cf. http://hul.harvard.edu/jhove/) as a quality check.
The DR should recommend file formats for the submission packages, e.g. GeoTif for remote reconnaissance data, or Seed/MiniSeed as the format for geodata, as used in GeoFon (http://www.gfz-potsdam.de/geofon/).
The kopal project specifies the Universal Object Format (UOF) for submission packages.
Examples and the specifications can be found at:
http://kopal.langzeitarchivierung.de/index_objektspezifikation.php.de

9.2 The digital repository identifies which characteristics of the digital objects are significant for information preservation.

In determining the scope of the characteristics to be preserved, a balance should be struck between the goals concerning the technical possibilities and the costs of long-term preservation on the one hand and the needs of the designated community/communities on the other hand.

It may be effective to obtain different representations of an information object in order to preserve as many characteristics as possible.

Regarding information from databases, it may be sufficient to archive the data as so-called "flat files" (including a precise description of the data structure).
With regard to electronic files, the DOMEA system specifies saving the individual documents as image files. This precludes full-text searches and the executability of some documents (Excel tables or PowerPoint presentations).
Regarding web pages containing text-image information, one archive can store only text information, a second, only images, and a third, the entire interrelation. The different objectives lead to correspondingly different archiving strategies.
Screenshots from a standard browser are taken of web pages, but the text information is also stored for ease of research.

Kunze, John: Future-Proofing The Web: What We Can Do Today, 2005. [43]

9.3 The digital repository has technical control of the digital objects in order to carry out long-term preservation measures.

Many digital objects contain technical features which restrict their use, either for commercial or legal reasons. For the long-term preservation of digital objects it is crucial that the DR is capable of opening and processing the objects with no restrictions. All technical restrictions on use must therefore

be removed before submission to the DR.

Internal settings may prevent e.g. copying, printing or saving of objects; other objects are encrypted and require the input of codewords or cannot be opened after expiry of a certain period or after a specified number of sessions.
"Music and publishing industry agree duplication of copy-protected works with German National Library", joint press information released by Deutsche Nationalbibliothek, Börsenverein des Deutschen Buchhandels and Bundesverbands der Phonographischen Wirtschaft, http://www.d-nb.de/wir/recht/vereinbarung.htm

10 Archival storage of the digital objects is undertaken to defined specifications.

At the heart of a digital repository is implementation of the actual archiving process. This covers definition of the archival packages, storage of the AIPs and implementation of the long-term preservation measures.

10.1 The digital repository defines its archival information packages (AIPs).

Archival information packages are conceptual units which consist of content data and all metadata required for long-term preservation (cf. 12).
The definition of the archival information packages should include determination of the package and object structures used, plus suitable storage locations and formats.
Selection of the archival information packages should depend on the object types (e.g. digital script or 3D animated clip) and the characteristics of the objects to be preserved.

Open, disclosed and frequently used formats are preferred as archive file formats, the assumption being that these will have a longer life, and there are more likely to be techniques and tools for converting or emulating them, given that they are supported by a wide circle of users.

Examples of currently used archive formats, although their suitability should be checked in each individual case:

- for unformatted text: ASCII/Unicode
- for structured text: XML (http://www.w3.org/XML/)
- for formatted text: PDF/A (ISO 19005-1: 2005, http://www.iso.org/iso/catalogue_detail?csnumber=38920
- for raster graphics: TIFF 6.0 (http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf)
- for audio formats: WAVE (http://msdn.microsoft.com/en-us/library/ms713498%28VS.85%29.aspx)
- for video files: MPEG 4 File Format (ISO/IEC 14496, http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm)
- for executable programs: the source text and documentation of the programming language

When making a decision regarding (lossless) compression of data, a balance should be struck between driving extra memory space on the one

hand and subsequent dependency on the compression method on the other. Open or disclosed methods are preferable.

For the structural description of the archival information packages, XML is currently favoured, especially the METS schema, which allows the metadata and references to be managed for the individual files of an object or the files themselves.

Coy, Wolfgang: nestor - materialien 5: Perspektiven der Langzeitarchivierung multimedialer Objekte 2006. [44]

Witthaut, Dirk; Zierer, Andrea et al.: nestor - materialien 2: Digitalisierung und Erhalt von Digitalisaten in deutschen Museen, 2005. [45]

Erpanet-Conference: "File Formats for Preservation", 2004. [46]

Abrams, Stephen: Digital Formats And Preservation, 2005. [47]

LOC: Sustainability of Digital Formats. Planning for Library of Congress Collections, 2006. [48]

ISO 19005-1:2005: Document management - Electronic document file format for long-term preservation, 2005. [49]

Helfer, Bernward; Lupprian, Karl-Ernst: Dateiformate: Eigenschaften und Eignung für die Archivierung elektronischer Unterlagen. Eine Handreichung für Archivarinnen und Archivare, 2004. [50]

BSI: IT-Grundschutz-Kataloge : Auswahl geeigneter Datenformate für die Archivierung von Dokumenten, 2007. [51]

Gutzmann, Ulrike; Kamp, Ulrich et al.: Praktische Lösungsansätze zur Archivierung digitaler Unterlagen: "Langzeitarchivierung" und dauerhafte Sicherung der digitalen Überlieferung, 2007. [52]

AK Elektronische Archivierung: Matrix zur Bewertung von Dateiformaten, 2006. [53]

Van Wijk, Caroline; Rog, Judith: Evaluating File Formats for Long-term Preservation, 2007. [54]

Steinke, Tobias: Universelles Objektformat: Ein Archiv- und Austauschformat für digitale Objekte, 2006. [55]

10.2 The digital repository takes care of transforming the submission information packages (SIPs) into archival information packages (AIPs).

As part of the ingest process, the SIPs should be transferred into AIPs and specific long-term preservation metadata added. This might involve conversion of the format.

DOC files are converted into PDF/A files.

10.3 The digital repository guarantees the storage and readability of the archival information packages (AIPs).

The DR should use appropriate methods to ensure that the archival information packages are correctly stored and can be read using means available within the system. Readability here refers to the capacity to read the storage media and the bit sequence.

See 6.2 regarding ensuring the integrity of archival information packages.

Possibilities for storing and ensuring readability include:

- Use of RAID systems
- Digital storage on suitable media such as tapes, records, CDs, DVDs
- Analogue storage on microfilm. The users either access analogue images on the film or digital data following prior redigitisation (conversion strategy)

cf. the long-term storage methods developed in the ARCHE project (http://www.landesarchiv-bw.de/web/46253).
.

10.4 The digital repository implements strategies for the long-term preservation of the archival information packages (AIPs).

The long-term preservation measures specified in point 8 should be implemented. A time (or occasion) should be defined when each archival information package should be checked for whether a long-term preservation step – e.g. migration or the provision of emulation software – must be taken. If necessary, the relevant measure should be carried out and documented (cf. 12.4).

Such a strategy could involve e.g. determining that a check is to be made in 2007 to decide whether the documents stored in PDF1.1 have to be migrated to PDF/A.

11 The digital repository permits usage of the digital objects based on defined criteria.

The usage purposes described under point 2 must be specified at the object level. The objects can be used by individuals but also by client systems. The search and access possibilities regarding the access packages should be defined. Each search should result in a clear response from the system. If the DR is part of a larger archive, the connections between digital and analogue objects which belong together must also be given. This applies in particular for the constituent parts of hybrid objects.
Access packages are the information units which users receive as a response to inquiries to the DR.

11.1 The digital repository defines its dissemination information packages (DIPs).

The DR should define its dissemination information packages (DIPs) dependent upon the designated community/communities and the archival information packages (AIPs). A precondition for this is determining the

reference application environment in which the objects can be used. An archival information package may be offered in different dissemination information packages, depending on the usage context. Use of the information represented by the digital objects in most cases does not mean access to the archival information packages themselves, rather the use of copies or derivatives (possibly in combination with other information) which help interpretability. This could be a technical description, additional application software or emulation software.

To exchange data with other digital repositories, or to migrate the data to a different technical infrastructure it is necessary to transform parts of, or the entire content of, the DR into a documented, standardised export format. The information can thus be preserved beyond the life of the DR itself (cf. 4.5).

```
Picture library: for use in the Web, low resolution files which can be
displayed by today's browsers are generated from the master images. High-
resolution files can be supplied electronically for reproduction
purposes.
cf. http://www.bsb-muenchen.de/karten/bilddatenb.htm
The kopal project specifies the Universal Object Format (UOF) for the
AIP.  Examples and the specifications can be found at:
http://kopal.langzeitarchivierung.de/index_objektspezifikation.php.de
```

11.2 The digital repository ensures transformation of archival information packages (AIPs) into dissemination information packages (DIPs).

The dissemination information packages should be derived from the archival information packages according to a defined procedure. Dissemination information packages can be held in the digital repository and, in the event of changed conditions, be regenerated, or created directly from the archival information packages (on the fly) as required.

```
Information should be stored on the conversion process (conversion
software, parameters) for the conversion of high resolution master images
into low-resolution access versions which can be displayed by standard
browsers.
```

12 The data management system is capable of providing the necessary digital repository functions.

Data management is a cross-sectional process which supports the core processes of a DR – ingest, archival storage and access – and also the planning and implementation of the preservation measures, while ensuring integrity and authenticity at all stages of processing. The scope of the data management system is dictated by the goals of the DR.

Data management must perform the following tasks:

- identification of the digital objects and their relationships is essential for administration of the objects
- the precondition for finding and accessing digital objects is a formal description of their content and structure
- ensuring interpretability and integrity, and planning and implementing preservation measures presumes technical description of the objects
- documentation of all changes to the digital objects is necessary to ensure authenticity of the data
- recording of all legal restrictions and their basis (laws, regulations, contracts, agreements) is necessary to ensure that legal requirements are observed throughout the processing

These tasks are currently fulfilled by the generation and storage of metadata. Metadata can be recorded in a structured manner in a metadata plan. Various metadata schemata have become established for different purposes (e.g. descriptive, structural, technical, administrative, legal metadata) and for different fields (e.g. archives, libraries, museums). Orientation to a national or international standard or subsequent use of a widespread metadata schema is often possible and makes sense with regard in particular to the sustainability of the data, but also for cooperation and data exchange between producers / suppliers, the DR and users. A metadata schema contains defined fields (data elements) in which the respective content is recorded. The result is a data structure which can be used both by humans and machines.

The DR should establish rules for filling the fields with content (e.g. use of controlled terminology). Different tools permit the automatic generation or extraction of metadata (e.g. JHOVE for technical metadata).

In this criteria catalogue, metadata is treated as part of the conceptual information units: transfer package, archival package and access package. These can be managed e.g. in databases and/or XML structures.

Bischoff, Frank M.: Metadata in preservation: selected papers from an ERPANET seminar at the Archives School Marburg, 2004. [56]
METS: METS Schema 1.7 Documentation, 2007. [57]
PREMIS Working Group: Data Dictionary for Preservation Metadata, Version 2.0 2008. [32].
LMER, 2005. [58]
nestor - Handbuch : Metadatenstandards im Bereich der digitalen Langzeitarchivierung, nationale und internationale Entwicklungen, 2008. [59]

NATLIB: Metadata Standards Framework – Preservation Metadata (Revised), 2003. [60]

12.1 The digital repository uniquely and persistently identifies its objects and their relationships.

A DR should use internal identifiers to manage the objects and their parts and relationships (part/totality, different variants, versions etc.), especially for unique assignment of the content data to the metadata (cf. 12.7).

The use of externally visible, standardised persistent identifiers should ensure reliable referencing and citability of the objects.
Use of a resolving service allows persistent identifiers to be incorporated in the URL address, thereby ensuring permanent access. This requires ongoing data management of the resolving service.
The DR undertakes to support the resolving service in maintaining the data.

Transferred from the world of printed materials to electronic media are:

- Signatures
- ISBN (International Standard Book Number) for monographs
- ISBN (International Standard Book Number) for periodicals
  - ISBN (http://www.ietf.org/rfc/rfc3187.txt) and ISSN (http://www.ietf.org/rfc/rfc3044.txt) are registered as URN namespaces.

For electronic media other systems are used, e.g.:

- Uniform Resource Names (URN, http://www.ietf.org/rfc/rfc1737.txt): An international Internet standard for unique, permanent identification of objects. National Bibliography Numbers (NBNs), for instance, are used in libraries. These are sub-namespaces of the URNs, e.g.
  - URN: urn:nbn:de:0008-20050117016
    URL: http://nbn-resolving.de/urn:nbn:de:0008-20050117016
- The handle system (HDL, http://www.handle.net/):
  The handle system is used to provide functions which permit the issue, administration and resolution of PIs in the form of handles, e.g.
  - HDL: 1721.1/30592
    URL: http://hdl.handle.net/1721.1/30592
- Digital Object Identifiers (DOI, http://www.doi.org/)
  DOIs are used by publishers, but also increasingly for specialist and primary data. Their technical basis is the handle system, e.g.
  - DOI: 10.1045/april2004-dobratz
    URL: http://dx.doi.org/10.1045/april2004-dobratz
- Archival Resource Keys (ARK, http://www.ietf.org/internet-drafts/draft-kunze-ark-10.txt) This identifier can be used in different ways: as a link between an object and the institution responsible, between an object and its metadata or to an object or its commensurate copy, e.g.
  - http://foobar.zaf.org/ark:/12025/654xz321/s3/f8.05v.tiff
- Scientific reference linking system (SRef, http://www.sref.org/site/index.php)

DNB: Persistent Identifier: Der Service der Deutschen Nationalbibliothek. [61]
PADI: Persistent Identifiers. [62]

Erpanet-Conference: "Persistent Identifier", 2004. [63]
nestor - Handbuch : Persistent Identifier, 2008. [64]

12.2 The digital repository records adequate metadata for formal and content-based description and identification of the digital objects.

The scope, structure and content of the descriptive metadata should depend on the goals of the DR, its designated community / communities and the object types. Formal and content-based description of the objects in the form of metadata makes it possible to find objects; this is essential for the search options which are offered to users.

A number of different schemata have become established in the different fields:
Libraries:
Dublin Core http://dublincore.org/,
Automated Library Exchange Format (MAB, http://www.d-nb.de/standardisierung/formate/mab.htm),
Machine-Readable Cataloging (MARC, http://www.loc.gov/marc/),
Metadata Objects Description Schema (MODS, http://www.loc.gov/standards/mods/).
Library codes can be used in combination with this, e.g. RAK or AACR2 for formal identification, and RSWK or a classification (e.g. DDC, RVK) for content identification.
Archives:
General International Standard Archival Description (ISAD(G) http://www.icacds.org.uk/eng/ISAD(G).pdf),
Encoded Archival Description (EAD, http://www.loc.gov/ead/), supplemented by Encoded Archival Context (EAC, http://jefferson.village.virginia.edu/eac/).
For space-related data: ISO Standard 19115.
NASA DIF (Data Interchange Format, http://gcmd.nasa.gov/User/difguide/difman.html) as NASA descriptive format which has developed into a de-facto standard, and which is also used for the Global Change Master Directory (http://gcmd.nasa.gov/).

Shepherd, Elizabeth; Smith, Charlotte: The Application of ISAD(G) to the Description of Archival Datasets, 2000. [65]
DOMEA : Organisationskonzept, 2005. [66]
GDA: Metadaten für die Aussonderung und Archivierung digitaler Sachakten, 2004. [67]

12.3 The digital repository records adequate metadata for structural description of the digital objects.

The structure of complex objects must be adequately described so that they can be reconstructed and used as entire entities.

A range of different standards - often based on XLM schemata - can be used for representing the structures of digital objects:
METS, TEI (Text Encoding Initiative, http://www.tei-c.org/), EAD.
However, structure information can also be managed in the descriptive metadata and in the metadata for long-term preservation (e.g. PREMIS and

LMER).
A digital record generally consists of procedures which in turn consist of documents to which further documents (appendices) may belong. This hierarchy is described by a file which contains metadata about each level, and at the document level contains metadata and references to the documents themselves (primary information).
The digitised version of a conventional book consists of 200 individual image files. The metadata should list the correct order of the book pages and the corresponding image files.
An archived website consists of a number of HTML pages and JPEG image files which are bound to each other via links. These links should be recorded in the metadata.

12.4 The digital repository records adequate metadata to document all the changes made by the digital repository to the digital objects.

The DR should document all changes made to the digital objects. This also includes recording the people and systems involved and the corresponding rights (cf. 3.2). This documents the authenticity (cf. 7) and also ensures technical preservation of the digital objects.
In particular, the digital objects in DRs which have chosen migration as their long-term preservation strategy are frequently changed. Added to this are the transformations which are carried out during submission to the DR and for delivery of access objects.

This metadata (history, audit trail, provenance) can be managed: e.g. by METS (amdSec digiorivMD section), PREMIS (Events section), LMER (Processes section).
An archive should migrate objects stored in an obsolete data format to a valid format using a conversion program. Metadata on the migration procedure, the technical protocol, the time of migration, the factors involved (staff and technical aids) and the result of the action should be recorded and saved.

12.5 The digital repository acquires adequate metadata for technical description of the digital objects.

To ensure interpretability and integrity and to control the preservation measures, the objects themselves and, in the case of complex objects, all their files must be comprehensively described in technical terms.
This includes in particular the description of the representation information.

The technical description contains general information which can be used for all file formats, including:

- File name, storage location
- File size, different check sums
- Full description of file formats
- Hardware/software environment used for generation
- Hardware/software environment required for use
- Recording of all necessary additional objects (DTD, schema file, fonts etc.)

- The software used during the transfer to the archive (e.g. offline browser)

Also there is specific information for the individual formats, e.g. resolution, colour space, compression etc. for TIFF files.

The general technical metadata is also managed by METS (amdSec, techMD sections), PREMIS or LMER.

Other standards have become established for format-specific metadata:

- for images: Metadata for Images in XML schema (MIX, http://www.loc.gov/standards/mix/]), based on NISO Technical Metadata for Digital Still Images
- for text: as extension of METS schema: textmd.xsd

File format registers can be referenced to describe file formats, e.g.:

Global Digital Format Registry: http://hul.harvard.edu/gdfr/,

PRONOM: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx.

Tools are available to identify file formats, e.g. Digital Record Object Identification (DROID, http://droid.sourceforge.net/wiki/index.php/Introduction) and to automatically extract technical metadata, e.g. JSTOR/Harvard Object Validation Environment (JHOVE, http://hul.harvard.edu/jhove/).

Technical metadata does not necessarily have to be explicitly generated and managed in all cases, it may be extracted only as required in some cases (e.g. before migration).

Some examples:

A DR stores files in version 1.4 PDF files. "Acrobat Reader 5.0" is required to view the files. This program runs on a Microsoft operating system - Windows 98 SE or later. The entire software, however, requires a PC with a processor of at least 350 MHz and 64 MB main memory. These technical details are part of the metadata which is recorded and stored by the DR.

A DR stores files in A-1 PDF files. This format is described in full in ISO standard 19005-1:2005. The DR appends the relevant ISO standard to the metadata or refers to it via a reference within the metadata.

A DR stores files in XML format. The relevant schema files are required to assess the validity of these files. The DR appends the relevant schema files to the metadata or refers to them via a reference within the metadata.

12.6 The digital repository acquires adequate metadata to record the corresponding usage rights and conditions.

Use of the digital objects may be restricted for legal or contractual reasons. Depending on these conditions and the corresponding user groups, these rights and conditions must be recorded in such as way as to permit use to be controlled (e.g. controlled access, anonymised user copies) and the users to be informed of them (cf. 3.3).

A DR archives databases which are only released for use after a period of 60 years. The exception is use for scientific research. The usage restriction is part of the DR's metadata and reference is also made to the relevant legal paragraphs (in this case Art. 2 paragraph 4 clause 2 and Article 5 paragraph 3 of the Federal Archive Act and Article 16 paragraphs 6-9 of the Federal Statistics Act).

This can be effected e.g. in the METS RightsDeclationMD Extension Schema.

For issuing of author-based rights with markup, e.g. Creative Commons (http://www.creativecommons.org) or as in DPPL (Digital Peer Publishing License), see http://www.dipp.nrw.de/lizenzen

12.7 The package structure is preserved at all times.

The connection between the metadata and the content data must be secure and unambiguous.

This can be achieved e.g. by:
a) the use of internal yet externally visible persistent identifiers for the digital objects and their parts, especially content data and metadata (cf. 12.1)
b) holding all content and metadata belonging to an object in a directory or a file.

```
The metadata schema METS offers the possibility of filing a digital
object which has been converted into a string of ASCII characters by
means of a base 64 converter within an XML document together with the
metadata.
```

C. Infrastructure and Security

Infrastructure and Security looks at the technical aspects of the overall system and aspects of security.

13 The IT infrastructure is appropriate.

The IT infrastructure should put the specifications for handling the objects into practice on the technology and security levels. It is responsible for the totality of all the objects.

13.1 The IT infrastructure implements the object management requirements.

The requirements specified by the DR regarding the handling of objects should be implemented by the overall system at all stages of processing. This includes the main processes (in OAIS: "functional entities") of ingest, archival storage (incl. preservation actions) and access and the data management support process. Extension of these functions may become necessary as a result of the DR's goals.

```
Web-Ingest-Module, module for bulk ingest in batch operation
Storage module with possibility to resort to a different, geographically
remote storage system.
Usage module
If the DR policy includes registered users being able to feed their photo
collections themselves into the DR, assuming these are available as JPEG
files, the DR must then provide a suitable upload interface for users.
```

Borghoff, Uwe M.: nestor - materialien 3: Vergleich bestehender Archivierungssysteme, 2005. [68]

13.2 The IT infrastructure implements the security requirements of the IT security system.

Object management security requirements should be taken into consideration during realisation:
Ensuring the **integrity** of the objects, i.e. protecting them from illegal modifications arising from deliberate and unintentional human actions, and technical imperfection
Ensuring the **authenticity** of the objects.
Ensuring the **confidentiality** of the objects, i.e. excluding the possibility of

unauthorised procurement of information
Ensuring the **availability** of the objects through availability of the object management functions (protection against sabotage, system failures etc.)

Access to protected data (e.g. archived STASI investigation committee documents) must be restricted to authorised users by means of appropriate technical security precautions (e.g. passwords or biometric access barriers).
The use of approved digital signatures as defined in the Digital Signature Act, and time stamps for the preservation of patent applications.

14 The infrastructure protects the digital repository and its digital objects.

The infrastructure should protect the digital objects from system-based and external hazards. System-based hazards could arise e.g. due to hardware problems or the failure of individual storage media. Externally the DR's first priority must be to protect against natural threats (e.g. fire, water, seismic activity), and also against risks caused by humans. The objects can be harmed directly by employees or through harmful programs smuggled into the system (e.g. viruses). Protecting the data also involves preventing the unintentional forwarding of information by programs (trojans) or people (espionage).

The protection should cover the objects, the facilities used by the DR, the hardware, software and, not least, the staff.

The different risks must be countered by a package of technical (e.g. virus protection programs) and organisational (e.g. access restrictions) measures.

A fire which breaks out in the main building of the institution housing the DR should not result in damage to the objects or data loss, as there should be a suitable backup system at a separate location which can assume operations in the event of an accident.

BSI: Leitfaden IT-Sicherheit. IT-Grundschutz kompakt, 2007. [69]
BSI: IT-Grundschutz-Kataloge, 2007. [70]

III. Checklist

It is not possible to provide an absolute assessment of the measures for fulfilling the criteria. The assessment is always based on the goals of the digital repository; however the adequacy of the measures should be checked.
Besides implementation of the criteria, publication of appropriate documentation helps increase the transparency of the digital repository, and confidence in it.
This is why the 4 phases of fulfilment (1. Conception, 2. Planning and specifications, 3. Realisation and implementation, 4. Evaluation) and also publication are to be taken into consideration.

A	Organisational framework
1	**The DR has defined its goals.**
1.1	The DR has developed criteria for the selection of its digital objects.
1.2	The DR assumes responsibility for long-term preservation of the information represented by the digital objects.
1.3	The DR has defined its designated community/communities.
2	**The DR grants its designated community/communities adequate access to the information represented by the digital objects.**
2.1	The DR ensures its designated community/communities can access the digital objects.
2.2	The DR ensures that the designated community/communities can interpret the digital objects.
3	**Legal and contractual rules are observed.**
3.1	Legal contracts exist between producers and the digital repository.
3.2	In carrying out its archiving tasks, the DR acts on the basis of legal arrangements.
3.3	With regard to use, the DR acts on the basis of legal arrangements.
4	**The organisational form is appropriate for the DR.**
4.1	Adequate financing of the digital repository is secured.
4.2	Sufficient numbers of appropriately qualified staff are available.
4.3	Appropriate organisational structures exist for the DR.
4.4	The DR engages in long-term planning.
4.5	The DR reacts to substantial changes.
4.6	Continuation of the preservation tasks is ensured even beyond the existence of the DR.
5	**The digital repository undertakes appropriate quality management.**
5.1	All processes and responsibilities have been defined.
5.2	The DR documents all its elements based on a defined process.

B Object management

6 The DR ensures the integrity of the digital objects during all processing stages.

6.1 Ingest: the DR ensures the integrity of the digital objects.

6.2 Archival Storage: the DR ensures the integrity of the digital objects.

6.3 Access: the DR ensures the integrity of the digital objects.

7 The DR ensures the authenticity of the digital objects during all processing stages.

7.1 Ingest: the DR ensures the authenticity of the digital objects.

7.2 Archival Storage: the DR ensures the authenticity of the digital objects.

7.3 Access: the DR ensures the authenticity of the digital objects.

8 The DR has a strategic plan for its technical preservation measures.

9 The DR accepts digital objects from the producers based on defined criteria.

9.1 The DR specifies its submission information packages (SIPs).

9.2 The DR identifies which characteristics of the digital objects are significant for information preservation.

9.3 The DR has technical control of the digital objects in order to carry out long-term preservation measures.

10 Archival storage of the digital objects is undertaken to defined specifications.

10.1 The DR defines its archival information packages (AIPs).

10.2 The DR takes care of transforming the submission information packages (SIPs) into archival information packages (AIPs).

10.3 The DR guarantees the storage and readability of the archival information packages (AIPs).

10.4 The DR implements strategies for the long-term preservation of the archival information packages (AIPs).

11 The DR permits usage of the digital objects based on defined criteria.

11.1 The DR defines its dissemination information packages (DIPs).

11.2 The DR ensures transformation of archival information packages (AIPs) into dissemination information packages (DIPs).

12 The data management system is capable of providing the necessary digital repository functions.

12.1 The DR uniquely and permanently identifies its objects and their relationships.

12.2 The DR records adequate metadata for formal and content-based description and identification of the digital objects.

12.3 The DR records adequate metadata for structural description of the digital objects.

12.4 The DR records adequate metadata to record all the changes made by the

digital repository to the digital objects.

12.5 The DR acquires adequate metadata for technical description of the digital objects.

12.6 The DR acquires adequate metadata to record the corresponding usage rights and conditions.

12.7 The package structure is preserved at all times.

C. Infrastructure and Security

13 The IT infrastructure is adequate.

13.1 The IT infrastructure implements the object management requirements.

13.2 The IT infrastructure implements the security requirements of the IT security system.

14 The infrastructure protects the digital repository and its digital objects.

IV. Glossary and abbreviations

Access: An OAIS functional entity consisting of the functions and processes which make the archived information accessible to the users.

Archival Information Package (AIP): Information unit stored in the DR, consisting of content data and metadata required for long-term preservation.

Archival Storage: An OAIS functional entity consisting of the functions and processes ensuring the storage and availability of the archival packages.

Authenticity: The object is genuine; it represents, what it claims to represent.

Availability: The data is available to the user at the required time.

Confidentiality: Protection from unauthorised divulgence of the data.

Data: Formalised representation of information which permits it to be interpreted, processed and exchanged.

Designated community: An identifiable group of potential users with specific interests and circumstances. It could be the general public or a group of specialist scientists, for instance. It can be heterogeneous and consist of different user groups.

Dissemination Information Package (DIP): Information unit derived from one or more AIPs and which a user receives as a response to an inquiry to the DR. An access package consists of the data representing the content and, where applicable, the information needed for interpretation (e.g. a csv format file and description of the data structure; a DOS program in source code and emulation software for the DOS operating system).

Digital object: Logically discrete unit of digital data. This could be a simple object consisting of a single file (e.g. a PDF document) or a complex object consisting of a number of different files (e.g. an electronic journal consisting of individual articles saved as files). Further data (metadata) may be added to the information representing the content (content data) which e.g. serves the formal and content description, the structural description, the interpretability or the long-term preservation of the content (cf. submission package, archival package, access package).

Digital repository (DR): An organisation (consisting of people and technical systems) which has assumed responsibility for the long-term preservation and long-term availability of digital data and its provision for a specified designated community.

"Long-term" here means lasting beyond technological changes (to hard and software) and also any changes to the designated community (e.g. for future generations, indefinitely).

Ingest: An OAIS functional entity consisting of the functions and processes which receive the transfer packages from the producer/supplier, transform them into archival packages and incorporate them into the archive.

Integrity: 1. Completeness of the digital objects, 2. Exclusion of unintended modifications as defined in the preservation rules. The yardsticks for integrity are the characteristics of a digital object which are defined as worthy of preservation. Metadata can be created at different times in the lifecycle of digital objects (during production, archiving or provision for use etc.).

Metadata: Data representing information about other data by describing e.g. its content, structure, composition, handling, origin etc. The term is used primarily in the digital field (e.g. Dublin Core Metadata), although e.g. title listings in library catalogues, archive catalogue entries etc. can also be regarded as metadata. Metadata should be seen as parts of the conceptual units of transfer, archival and access packages.

OAIS: Reference model (ISO 14721:2003) for DRs which describes the core processes of a DR (functional entities) and provides an information model.

Preservation planning (long-term preservation measures): The totality of all methods specifically used to archive digital objects indefinitely and to make them available over a sustained period. This includes methods for the physical preservation of the data and also the use of migration and emulation techniques to change the archived objects or their environments to guarantee their future use.

Producer: People or client systems who/which transfer digital objects to the DR for long-term preservation. They are not necessarily the originators; they could also be the suppliers of the digital objects.

Quality: The quality of a DR is the degree to which a number of inherent characterising properties fulfil the specified requirements. Requirements here are prerequisites or expectations which are laid down and which are generally taken for granted or compulsory (following ISO 9000:2000).

Representation information: Information which is necessary to interpret digital data (e.g. the file format of a file).

Submission Information Package (SIP): Information unit submitted by the producer to the DR. The content data may already be supplemented with metadata.

Trustworthiness: Trustworthiness is the capacity of a system to operate in accordance with its objectives and specifications (i.e. it does exactly what it claims to do). The trustworthiness of a DR can be tested and assessed on the basis of a criteria catalogue.

Users: People or client systems who/which interact with the DR to find and use the information represented by the digital packages.

V. Bibliography

[1] CCSDS, Consultative Comitee for Space Data Systems: A Reference Model for an Open Archival Information System, Blue Book, 2002, http://public.ccsds.org/publications/archive/650x0b1.pdf, (28.11.2008).
Entwurf zu ISO 14721:2003, International Organization for Standardization: Space data and information transfer systems -- Open archival information system -- Reference model, 2003
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683 , (25.02.2010).

[2] DINI, Deutsche Initiative für Netzwerkinformation e.V., Arbeitsgruppe „Elektronisches Publizieren“ DINI – Zertifikat. Dokumenten- und Publikationsserver 2007, Version 2.0, 2006, http://edoc.hu-berlin.de/series/dini-schriften/2006-3/PDF/3.pdf (25.02.2010).
[english: DINI-Certificate Document and Publication Services 2007 http://edoc.hu-berlin.de/series/dini-schriften/2006-3-en/PDF/3-en.pdf (14.01.2010)]

[3] RLG, Working Group on Digital Archive Attributes: Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report, Mountain View CA, 2002, http://www.oclc.org/programs/ourwork/past/trustedrep/repositories.pdf, (25.02.2010).

[4] RLG – NARA Task Force on Digital Repository Certification: Audit Checklist for Certifying Digital Repositories. Draft for Public Comment, 2005, http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2416.pdf, (25.02.2010).
CRL, OCLC/RLG – NARA Task Force on Digital Repository Certification: Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), 2007, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf, (25.02.2010).

[5] DRAMBORA, Digital Repository Audit Method Based on Risk Assessment, http://www.repositoryaudit.eu/, (25.02.2010).

[6] DCC, Digital Curation Centre, http://www.dcc.ac.uk/, (25.02.2010).

[7] DPE, Digital Preservation Europe, http://www.digitalpreservationeurope.eu/, (25.02.2010).

[8] CRL, Centre for Research Libraries: Certification of Digital Archives Project, http://www.crl.edu/archiving-preservation/digital-archives/past-projects/cda,

(25.02.2010).

CRL, Centre for Research Libraries: Long-Lived Digital Collections http://www.crl.edu/archiving-preservation/digital-archives/long-lived-digital-collections, (25.02.2010).

[9] DCC, DPE, nestor, CRL: Core Requirements for Digital Archives, Ten Principles, Chicago, 2007, http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re, (25.02.2010).

[10] Erpanet-Conference, Electronic Resource Preservation and Access Network: "Policies for Digital Preservation", 2003, http://www.erpanet.org/events/2003/paris/ERPAtraining-Paris_Report.pdf, (25.02.2010).

[11] Erpanet-Conference, Electronic Resource Preservation and Access Network: "Appraisal of Scientific Data", 2003, http://www.erpanet.org/events/2003/lisbon/LisbonReportFinal.pdf, (25.02.2010).

[12] Wiesenmüller, Heidrun; Jendral, Lars et al.: Auswahlkriterien für das Sammeln von Netzpublikationen im Rahmen des elektronischen Pflichtexemplars: Empfehlungen der Arbeitsgemeinschaft der Regionalbibliotheken. In: Bibliotheksdienst (2004), No 11, S. 1423-1444, http://www.zlb.de/aktivitaeten/bd_neu/heftinhalte/heft9-1204/digitalebib1104.pdf, (25.02.2010).

[13] BArchG, Bundesarchivgesetz: Gesetz über die Sicherung und Nutzung von Archivgut des Bundes, Koblenz, 2005, http://www.bundesarchiv.de/benutzung/rechtsgrundlagen/bundesarchivgesetz/index.html, (25.02.2010).

[14] Goebel, Jürgen W.; Scheller, Jürgen et al.: nestor - materialien 1: Digitale Langzeitarchivierung und Recht, Frankfurt am Main, 2004, urn:nbn:de:0008-20040916022, http://files.d-nb.de/nestor/materialien/nestor_mat_01.pdf, (25.02.2010).

[15] Coyle, Karen: Rights in the PREMIS Data Model. A Report for the Library of Congress, Washington, 2006, http://www.loc.gov/standards/premis/Rights-in-the-PREMIS-Data-Model.pdf, (25.02.2010).

[16] Erpanet-Conference, Electronic Resource Preservation and Access Network: "Business Models related to Digital Preservation", 2004, http://www.erpanet.org/events/2004/amsterdam/Amsterdam_Report.pdf, (25.02.2010).

[17] Digital Longevity Department: Vers van de pers...Kostenmodel digitale bewaring (Kostenmodell für die Langzeitarchivierung), Den Haag, 2006, http://www.digitaleduurzaamheid.nl/detail.cfm?id=106&sub=nieuws&categorie=0, (25.02.2010).

[18] Palm, Jonas: The Digital Black Hole, Stockholm, 2006, http://www.tape-online.net/docs/Palm_Black_Hole.pdf, (25.02.2010).

[19] Oltmans, Erik; Kol, Nanda: A Comparison between Migration and Emulation in Terms of Costs. In: RLG diginews (2005), Vol 9, No 2, http://worldcat.org/arcviewer/1/OCC/2007/07/10/0000068902/viewer/file1.html#article0, (25.02.2010).

[20] ISO 9000:2005, International Organization for Standardization: Quality management systems. Fundamentals and vocabulary, 2005, http://www.iso.org/iso/catalogue_detail?csnumber=42180, (25.02.2010).

[21] Liggesmeyer, Peter: Software-Qualität. Testen, Analysieren und Verifizieren von Software, Heidelberg/Berlin, 2002.

[22] Kneuper, Ralf: CMMI. Verbesserung von Softwareprozessen mit Capability Maturity Model Integration, 2. Auflage, Heidelberg, 2006.

[23] ITIL: IT - Infrastructure Library, http://www.itil.org/, (25.02.2010).

[24] CCSDS, Consultative Comitee for Space Data Systems: Producer-Archive Interface Methodology - Abstract Standard, Blue Book, Washington, 2004, http://public.ccsds.org/publications/archive/651x0b1.pdf, 22.09.2008. Entwurf zu ISO 20652:2006, International Organization for Standardization: Space data and information transfer systems -- Producer-archive interface -- Methodology abstract standard, 2006, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39577, (25.02.2010).

[25] Erpanet-Conference, Electronic Resource Preservation and Access Network: "Workshop on Workflow", 2004, http://www.erpanet.org/presentations.php, (25.02.2010).

[26] Borghoff, Uwe M.; Rödig, Peter et al.: Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente, Heidelberg, 2003. [english: Borghoff, Uwe M.; Rödig, Peter et al.: Long-term preservation of digital documents. Principles and practices, Berlin, 2006.]

[27] ISO 15489-1:2001, International Organization for Standardization: Information and documentation - Records Management, 2001, http://www.iso.org/iso/catalogue_detail?csnumber=31908, (25.02.2010).

[28] Shirey, Robert W.: Internet Security Glossary, Version2, 2007, http://tools.ietf.org/html/rfc4949, (25.02.2010).

[29] ISO/IEC 15408, International Organization for Standardization: Information technology. Security techniques - Evaluation criteria for IT security, Part 1, 2005; Part 2, 2008, Part 3, 2008, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40612, (25.02.2010)

[30] DigiCult, Digital Culture: Integrity and Authenticity of Digital Cultural Heritage Objects, 2002, http://www.digicult.info/downloads/thematic_issue_1_final.pdf, (25.02.2010).

[31] Littman, Justin: A Technical Approach and Distributed Model for Validation of Digital Objects. In: D-Lib Magazine (2006), Vol 12, No 5, http://www.dlib.org/dlib/may06/littman/05littman.html, (25.02.2010).

[32] PREMIS Working Group, Preservation Metadata: Implementation Strategies: Data Dictionary for Preservation Metadata, Version 2.0, Washington, 2008, http://www.loc.gov/standards/premis/v2/premis-2-0.pdf, (25.02.2010).

[33] Gladney, Henry M.; Bennett, John L.: What Do We Mean by Authentic? What's the Real McCoy? In: D-Lib Magazine (2003), Vol 9, No 7/8, http://www.dlib.org/dlib/july03/gladney/07gladney.html, (25.02.2010).

[34] InterPARES, The International Research on Permanent Authentic Records in Electronic Systems, http://www.interpares.org/, (25.02.2010).

[35] nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland: Digitale Erhaltungsstrategien. In: nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.2, 2008, Kapitel 12, http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf, (25.02.2010).

[36] Rauch, Carl; Rauber, Andreas: Anwendung der Nutzwertanalyse zur Bewertung von Strategien zur langfristigen Erhaltung digitale Objekte, Wien, 2006, http://www.ifs.tuwien.ac.at/~andi/publications/pdf/rau_zfbb05.pdf, (25.02.2010).

[37] Strodl, Stephan; Becker, Christoph et al.: How to Choose a Digital Preservation Strategy: Evaluating a Preservation Planning Procedure, Wien, 2007, http://www.ifs.tuwien.ac.at/~strodl/paper/FP060-strodl.pdf, (25.02.2010).

[38] DOMEA, Dokumentenmanagement und elektronische Archivierung: Aussonderung und Archivierung elektronischer Akten. Erweiterungsmodul zum DOMEA-Organisationskonzept 2.1, 2005,

http://www.verwaltung-innovativ.de/nn_684674/SharedDocs/Publikationen/DE/domea__konzept__aussonderung__und__archivierung__elektronischer__akten,templateId=raw,property=publicationFile.pdf/domea_konzept_aussonderung_und_archivierung_elektronischer_akten.pdf, (25.02.2010).

[39] The U.S. National Archives & Records Administration: Disposition of Federal Records. Subpart L -- Transfer of Records to the National Archives of the United States, Part 1228, § 1228.270. In: NARA Regulations (2002), http://www.archives.gov/records-mgmt/pdf/dfr-2000.pdf, (25.02.2010).

[40] NDAD, National Digital Archive of Datasets: Transfer Procedures (Overview), 2005, http://www.ndad.nationalarchives.gov.uk/resources/pdf/xfer_notes_overview.pdf, (25.02.2010).

[41] DPC, Digital Preservation Coalition: Decision Tree for Selection of Digital Materials for Long-term Retention, York, 2006, http://www.dpconline.org/technology-watch-reports/download-document/298-preservation-handbook-decision-tree.html?q=decision+tree, (25.02.2010).

[42] nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland: Wege ins Archiv: Ein Leitfaden für die Informationsübernahme in das digitale Langzeitarchiv, 2008, urn:nbn:de:0008-2008103009, http://files.d-nb.de/nestor/materialien/nestor_mat_10.pdf (25.02.2010). nestor - the german Network of expertise in Digital long-term preservation: Into the Archive - a guide to the information transfer to a digital repositiory, 2009, http://files.d-nb.de/nestor/materialien/nestor_mat_10_en.pdf, (25.02.2010).

[43] Kunze, John: Future-Proofing The Web: What We Can Do Today. In: iPRES - International Conference on Digital Preservation (2005), http://rdd.sub.uni-goettingen.de/conferences/ipres05/download/Future-Proofing%20The%20Web%20What%20We%20Can%20Do%20Today%20-%20John%20Kunze.pdf, (25.02.2010).

[44] Coy, Wolfgang: nestor - materialien 5: Perspektiven der Langzeitarchivierung multimedialer Objekte, Frankfurt am Main, 2006, urn:nbn:de:0008-20051214015, http://edoc.hu-berlin.de/series/nestor-materialien/5/PDF/5.pdf, (25.02.2010).

[45] Witthaut, Dirk; Zierer, Andrea et al.: nestor - materialien 2: Digitalisierung und Erhalt von Digitalisaten in deutschen Museen, Frankfurt am Main, 2005,

urn:nbn:de:0008-20041223022,
http://files.d-nb.de/nestor/materialien/nestor_mat_02.pdf, (25.02.2010).

[46] Erpanet-Conference, Electronic Resource Preservation and Access Network: "File Formats for Preservation", 2004,
http://www.erpanet.org/presentations.php, (25.02.2010).

[47] Abrams, Stephen: Digital Formats And Preservation. In: iPRES - International Conference on Digital Preservation (2005),
http://rdd.sub.uni-goettingen.de/conferences/ipres05/download/Digital%20Formats%20And%20Preservation%20-%20Stephen%20Abrams.pdf, (25.02.2010).

[48] LOC, Library of Congress: Sustainability of Digital Formats. Planning for Library of Congress Collections, 2006,
http://www.digitalpreservation.gov/formats/index.shtml, (25.02.2010).

[49] ISO 19005-1:2005: Document management - Electronic document file format for long-term preservation, Part 1: Use of PDF 1.4 (PDF/A-1), 2005.
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920 (25.02.2010).

[50] Helfer, Bernward; Lupprian, Karl-Ernst: Dateiformate: Eigenschaften und Eignung für die Archivierung elektronischer Unterlagen. Eine Handreichung für Archivarinnen und Archivare, Wiesbaden und München, 2004,
http://www.gda.bayern.de/datfor.pdf, (25.02.2010).

[51] BSI, Bundesamt für Sicherheit in der Informationstechnik: Auswahl geeigneter Datenformate für die Archivierung von Dokumenten. In: IT-Grundschutz-Kataloge, 10. Ergänzungslieferung, 2008,
https://www.bsi.bund.de/cln_183/ContentBSI/grundschutz/kataloge/m/m04/m04170.html , (14.01.2010).
[english:
https://www.bsi.bund.de/cln_183/EN/topics/ITGrundschutz/ITGrundschutzCatalogues/itgrundschutzcatalogues_node.html (14.01.2010)]

[52] Gutzmann, Ulrike; Kamp, Ulrich et al.: Praktische Lösungsansätze zur Archivierung digitaler Unterlagen: "Langzeitarchivierung" und dauerhafte Sicherung der digitalen Überlieferung. In: Archiv und Wirtschaft (2007), No 1,
http://www.wirtschaftsarchive.de/akea/handreichung.htm, (25.02.2010).

[53] AK Elektronische Archivierung, VdW Arbeitskreis "Elektronische Archivierung": Matrix zur Bewertung von Dateiformaten, Frankfurt am Main, 2006,
http://www.wirtschaftsarchive.de/akea/Dateiformate_Bewertung_V0.1.xls,
(25.02.2010)..

[54] Van Wijk, Caroline; Rog, Judith: Evaluating File Formats for Long-term Preservation. In: iPRES – International Conference on Digital Preservation (2007), http://rdd.sub.uni-goettingen.de/conferences/ipres07/presentations/Caroline-iPRES2007-11-12oct_CW.pdf, (25.02.2010).

[55] Steinke, Tobias: Universelles Objektformat: Ein Archiv- und Austauschformat für digitale Objekte, Frankfurt am Main, 2006, http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf, (25.02.2010).
[english: Universal Object Format. An archiving and exchange format for digital objects, http://kopal.langzeitarchivierung.de/downloads/kopal_Universal_Object_Format.pdf, (30.10.2009).]

[56] Bischoff, Frank M.: Metadata in preservation: selected papers from an ERPANET seminar at the Archives School Marburg. In: ERPANET Seminar: "Metadata in Digital Preservation" (2004), http://www.erpanet.org/presentations.php, (25.02.2010).

[57] METS, Metadata Encoding and Transmission Standard: METS Schema 1.7 Documentation, 2007, http://www.loc.gov/standards/mets/mets-schemadocs.html, (25.02.2010).

[58] LMER, Langzeitarchivierungsmetadaten für elektronische Ressourcen, Frankfurt am Main, 2005, http://www.d-nb.de/standards/pdf/lmer12.pdf, (25.02.2010).
[english: LMER – Long-term preservation Metadata for Electronic Resources, http://www.d-nb.de/standards/pdf/lmer12_e.pdf, (30.10.2009)]

[59] nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland: Metadatenstandards im Bereich der digitalen Langzeitarchivierung, nationale und internationale Entwicklungen. In: nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.2, 2008, Kapitel 10.1, http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf, (25.02.2010).

[60] NATLIB, National Library of New Zealand: Metadata Standards Framework – Preservation Metadata (Revised), Wellington, 2003, http://www.natlib.govt.nz/downloads/metaschema-revised.pdf, (25.02.2010).

[61] DNB, Deutsche Nationalbibliothek, Persistent Identifier: Der Service der Deutschen Nationalbibliothek, http://www.persistent-identifier.de/, (25.02.2010).

[62] PADI, Preserving Access to Digital Information: Persistent Identifiers, http://www.nla.gov.au/padi/topics/36.html, (25.02.2010).

[63] Erpanet-Conference, Electronic Resource Preservation and Access Network: "Persistent Identifier", 2004, http://www.erpanet.org/presentations.php, (25.02.2010).

[64] nestor - Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland: Persistent Identifier. In: nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.2, 2008, Kapitel 13.2, http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf, (25.02.2010).

[65] Shepherd, Elizabeth; Smith, Charlotte: The Application of ISAD(G) to the Description of Archival Datasets. In: Journal for the Society of Archivists (2000), Vol. 21, No 1, S. 55-86.

[66] DOMEA, Dokumentenmanagement und elektronische Archivierung : Organisationskonzept, Version 2.1, 2005, http://www.verwaltung-innovativ.de/cln_117/SharedDocs/Publikationen/DE/domea__konzept__organisationskonzept__2__1,templateId=raw,property=publicationFile.pdf/domea_konzept_organisationskonzept_2_1.pdf, (25.02.2010).

[67] GDA, Generaldirektion der Staatlichen Archive Bayerns, Metadaten für die Aussonderung und Archivierung digitaler Sachakten München, 2004, http://www.gda.bayern.de/metadat.pdf, (25.02.2010).

[68] Borghoff, Uwe M.: nestor - materialien 3: Vergleich bestehender Archivierungssysteme, Frankfurt am Main, 2005, urn:nbn:de:0008-20050117016, http://files.d-nb.de/nestor/materialien/nestor_mat_03.pdf, (25.02.2010).

[69] BSI, Bundesamt für Sicherheit in der Informationstechnik: Leitfaden IT-Sicherheit. IT-Grundschutz kompakt, Bonn, 2007, https://www.bsi.bund.de/cln_183/DE/Themen/ITGrundschutz/LeitfadenInformationssicherheit/leitfaden_node.html, (14.01.2010).
[english: IT-Security Guidelines https://www.bsi.bund.de/cln_183/ContentBSI/EN/topics/ITGrundschutz/ITSecurityGuidelines/guidelines.html, (14.01.2010)]

[70] BSI, Bundesamt für Sicherheit in der Informationstechnik: IT-Grundschutz-Kataloge, 10. Ergänzungslieferung, 2008, https://www.bsi.bund.de/cln_183/ContentBSI/grundschutz/kataloge/kataloge.html, (14.01.2010).
[english: IT Grundschutz Catalogues https://www.bsi.bund.de/cln_183/EN/topics/ITGrundschutz/ITGrundschutzCatalogues/itgrundschutzcatalogues_node.html, (14.01.2010)]

Appendix

Core Requirements for Digital Archives [9]

These general criteria constitute the basis for further cooperation and collaboration.

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.

1. The repository commits to continuing maintenance of digital objects for identified community/communities.
2. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfill its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
7. 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.
8. Fulfills requisite dissemination requirements.
9. Has a strategic program for preservation planning and action.
10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.