Data Service Infrastructure for the Social Sciences and Humanities

EC FP7

Grant Agreement Number: 283646

Deliverable Report

Deliverable: D4.1
Deliverable Name: 4.1 Roadmap for Preservation and Curation in the SSH
Deadline: 31 December 2012
Nature: R

Responsible: NSD, CentERData
Work Package Leader: NSD

Contributing Partners and Editors: Vigdis Kvalheim (NSD), Dag Kiberg (NSD), Trond Kvamme (NSD), Eric Balster (CenterData), Marika de Bruijne (CenterData), Arnaud Wijnant (CenterData), Astrid Recker (GESIS), Przemek Lenkiewicz (MPI), Sally Widdop (CITY), Bartholomäus Wloka (OEAW).

Executive Summary

This report is a combination of a state-of-the-art description and a guideline for data repositories seeking to provide a trustworthy environment for data management and data curation within the social sciences and humanities. It aims to work as an implementable resource for existing and emerging repositories that seeks to provide trustworthy long-term data repository services. The report discusses the benchmark frameworks and standards for trusted digital repositories by showing how the different tools and standards can be implemented through a five-step trust maturity model, from initial self-assessment to external audit and formal certification. Content of the report is structured as follows: in the introduction we clarify some of the terminology in the report, before some general aspects and issues in long-term preservation is laid out. After the short descriptions of the assessed tools and standards in part 2, the five steps of the trust maturity model is presented in more detail in part 3, providing descriptions of guidelines and pointers to supporting tools and sources within each level. The level of abstraction is generally high; the processes describes are relevant for most information processing entities, not just data repositories within the SSH community.

Content

Terminology

Unless otherwise is pointed out, the definitions of terms below are taken from the OAIS (Open Archival Information System) reference model (CCSDS 650.0-M-2, 2012).

Archival Information Package (AIP): AIP, as defined in OAIS reference model is an information package that is used to transmit archival objects into a digital archival system, store the objects within the system, and transmit objects from the system. An AIP contains both metadata that describes the structure and content of an archived essence and the actual essence itself. It consists of multiple data files that hold either a logically or physically packaged entity. An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

Authenticity: The degree to which a person (or system) regards an object as what it is purported to be. Authenticity is judged on the basis of evidence.

Data Object: Either a Physical Object or a Digital Object.

Descriptive Information: The set of information, consisting primarily of Package Descriptions, which is provided to Data Management to support the finding, ordering, and retrieving of OAIS information holdings by Consumers.

Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the archive and this definition may change over time.

Dissemination Information Package (DIP): An Information Package, derived from one or more AIPs, and sent by Archives to the Consumer in response to a request to the OAIS.

Independently Understandable: A characteristic of information that is sufficiently complete to allow it to be interpreted, understood and used by the Designated Community without having to resort to special resources not widely available, including named individuals.

Information Object: A Data Object together with its Representation Information.

Information Package: A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.

Knowledge Base: A set of information, incorporated by a person or system that allows that person or system to understand received information.

Long Term Preservation: The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term.

Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in an OAIS. This period extends into the indefinite future.

Repository: In this context we use the term somewhat abstractly, to refer to a collection of services involved with the acquisition, management, and dissemination of digital material (DPE, 2008).

Representation Information: The information that maps a Data Object into more meaningful concepts.

Submission Information Package (SIP): An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.

Trusted Digital Repository: A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future (RLG-OCLC, 2002). A repository is trusted if it can demonstrate its capacity to fulfill its specified functions, and if those specified functions satisfy an agreed set of minimal criteria which all Trusted Repositories are assumed to require. At the very basic level, the definition of a trustworthy digital repository must start with 'a mission to provide reliable, long-term access to managed digital resources to its Designated Community, now and into the future (Giaretta, 2011).

1. Introduction

> *"Preservation was once an obscure background operation of interest chiefly to conservators and archivists. It is now widely recognised as one of the most important elements of a functional cyberinfrastructure."*
>
> - Our Cultural Commonwealth (ACLS, 2006)

1.1 Issues in Long Term Data Preservation

Massive increases in the availability of data, the 'flood of digital objects' within the social sciences and humanities are making progress possible in analyzing, understanding, and addressing many societal issues. These same forces also pose challenges to the scientific infrastructure supporting data sharing and data management (King, 2011).

One of the most important challenges is the growing interdependence among scientific disciplines; it has become more common in current sciences to take advantage of data captured in other scientific disciplines (Anderson, 2004). This has for a long time been true for disciplines within the natural sciences, and there is an increasing trend of 'cross-fertilization' of data within the SSH community to incorporate and take advantage of data from other scientific disciplines.

A second challenge is that the requirements for scientific data management are changing in response to the continuing developments of technology. The increasing use of and dependence on, computer-based technology is changing both the practices and products of science; digital information within the SSH communities are being produced by the changes and progress in computer technology and the many commercial entities creating and monetizing new forms of data repositories (King, 2011).

All research disciplines are faced with great challenges with respect to data creation, management, curation and access. The CORDIS project points to some of the most important issues (HLEG, 2010):

- How will we preserve the data? What will be the point of storing all this scientific data if, a century from now, it has degraded, been corrupted, or is simply too difficult for anyone but a well-equipped expert to use? As such, non-maintainability of essential hardware, software or support environment may make the information inaccessible and/or users may become unable to understand or use the data.

- How will we protect the integrity of the data? As the 'data tide rises' higher, how will we detect unauthorised alterations? Should every researcher, and indeed every citizen, have access to the data repositories? Should there be different levels of access allowed?

- How will we convey the context and provenance of the data? Given the emerging trend to make all publicly funded research data publicly available, just how will users from a wide range of backgrounds understand and query the data they are accessing, and recognise the special circumstances under which it was collected?

- What new funding and business models will we need, so that everyone – researchers, enterprises, citizens – have adequate incentive to contribute to the data infrastructure? What kinds of data, under what circumstances, should be free?

- How will we protect the privacy of individuals linked to the data on the one hand, while providing researchers access to vital data on the other hand?

As a result there is a growing awareness at the political level across Europe as well as within scientific communities that archiving and preservation of scientific data should as such not be thought of as a reactive activity. Formally agreed practices and procedures should be part of proactive data preservation routines in the SSH data curation community.

This is very much in line with the ideas behind the OECD Principles and Guidelines for Access to Research Data from Public Funding (OECD, 2007): "Specialised support services…should be considered as means to ensure the cost-effective production, use, management and archiving of research data", and argues that "with data management becoming ever more complex in certain areas of research, traditional informal arrangements between researchers may no longer be adequate and may need to be complemented by formally agreed practices and procedures."

This is a very important and message at a time when researchers, research projects, research groups, and institutions increasingly argue in favor of decentralized data management and even project based data deposit and dissemination services, pointing to the fact that hardware and tools are available at relatively low cost.

Many of these issues involve trust and give rise to questions regarding the trustworthiness of the stored information. Producers and consumers of information are questioning which "memory organizations" are capable of ensuring the authenticity, integrity, confidentiality and availability of digital information (NESTOR, 2009). The fundamental question is whether we can find authentic and trustworthy ways of preserving, protecting and providing continuously access to this information.

1.2 Long Term Data Preservation as Communication with the Future

Digital preservation can be thought of as "communication with the future" (Mois, Klas, & Hemmje, 2009) and as an integral part of the "memory of the world" (UNESCO, 2012). At the base of digital preservation lies the notion that information that is understood today must be transmitted to an unknown system in the future where it will be interpreted and displayed (Mois, Klas, & Hemmje, 2009). To ensure that the information can be understood by the consumers of the future information system, more than just the preservation of the bit stream has to be undertaken. The information needs to be maintained actively, through carefully planned and managed procedures; the aggregated bit stream, the data, has to be curated and 'value-added' to remain interpretable for future users (Mois, Klas, & Hemmje, 2009). One of the most important tasks in this context is to ensure that not only the data sent into the future remains its integrity and authenticity, but that the integrity and authenticity is also guaranteed for the whole preservation system and its collections (Schott, Dittmann, Vielhauer, Krätzer, & Lang, 2008). Hence, when building information systems, or what we in this context can call '*digital repositories*', the element of trust is of integral importance.

1.3 How to become a Trusted Digital Repository?

A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its *designated community*, now and in the future (RLG-OCLC, 2002). A repository is considered *trusted* if it can demonstrate its capacity to fulfill its specified functions, and if those specified functions satisfy an agreed set of minimal criteria which all *trusted repositories* are assumed to require. The requirement that the repository have to demonstrate compliance with these criteria is critical, with the result that the acquisition of trust is assumed to be largely synonymous with a defined set of *requirements specifications*, or through processes of *audit* and *certification* (DPE, 2008).

To this end, several initiatives have developed tools to enable repositories to be audited or self-assessed. These have been characterized by two complementary approaches. The TRAC (RLG-OCLC, 2002) (CRL, 2007) and NESTOR groups (NESTOR, 2009) have produced checklists of specific criteria which repositories are required to be able to fulfill and document in order to obtain certification. The work of the TRAC group later evolved into an international official standard (CCSDS 652.0-M-1, 2011). The checklist approach is concrete and specific and is well-suited for an external certification process. On the other hand it is somewhat strict and rigid and may be difficult to apply among the wide variety of digital repositories that might wish to seek trusted status.

Consequently, other projects like the PLATTER (DPE, 2008), DRAMBORA (DCC & DPE, 2007) and Data Seal of Approval (DSA, 2010) have developed toolkits that seek to guide repositories through a risk-assessment exercise which can enable them to evaluate, through self-assessment, their ability to fulfill their self-specified goals. These kinds of toolkits are more flexible because they assess a repository *relative* to the repository's own self-defined goals, not

to an externally defined standard. However, this implies that the trustworthiness of a self-assessed repository can only be as good as the fitness of those self-defined goals (DPE, 2008).

A suitable compromise between the detailed checklist-for-certification approach and the self-assessment approach would be to provide a guide to repositories which allows them to navigate their development of trust maturity through a step-wise process, from a low trust maturity level that can be improved through self-assessment, to higher maturity levels for process optimization and external repository auditing.

1.4 Scope and Aim of Report

The overall aim of the DASISH project is to provide solutions to a number of common infrastructure issues relevant for the five ESFRI projects in social science and humanities, being CESSDA, CLARIN, DARIAH, ESS and SHARE. DASISH has identified four major infrastructure issues of interest, namely data quality, data archiving, data access and legal and ethics issues. Through DASISH the aim is that the participating infrastructures will not only obtain new solutions for specific problems and a consolidation of their infrastructure building, but will work out solutions facilitating interdisciplinary cross-walks of their researchers. This will be of mutual benefit for the five infrastructures and the communities they serve.

The aim of Work Package 4 (Data Archiving) is to discuss the general state of long term data preservation in the Social Science and Humanities (SSH) domain and work out suggestions to overcome current bottlenecks and to establish trust in procedures. Based on this, WP4 aims to develop models for a robust deposit and long term preservation services which can be offered to all researchers within the SSH domain, including policy rules, and business and access models.

Deliverable 4.1 aims to provide a state-of-the-art report, with guidelines or recommendations for proper data management that can work as a set of requirement specifications for data preservation and curation. As such, this deliverable is aimed at scientific stakeholders who are likely to be interested in development of new data repositories or in the improvement of existing ones. It must also be seen as input to the assessment of existing institutional and academic deposit services that is part of deliverable 4.2, as well as the drafting of a comprehensive set of policy rules for data management in deliverable 4.4.

A well-defined set of guidelines for digital curation is crucial to the continued viability and trust of digital materials. This report provides an assessment of existing data repository models, or frameworks that provides checklists or guidelines to data preservation entities. Based on the assessment of the various models we aim to formulate a “guideline of guidelines” that can be used for a broad spectrum of digital long-term repositories and that aim to retain their validity over a longer period. Relatively abstract criteria have therefore been chosen.

As such, the report provides pointers to sources of advice and guidance and synthesizes the essence of existing guidelines concerning creation and management of digital materials. A well-defined set of guidelines for digital curation is crucial to the continued viability and trust of digital materials.

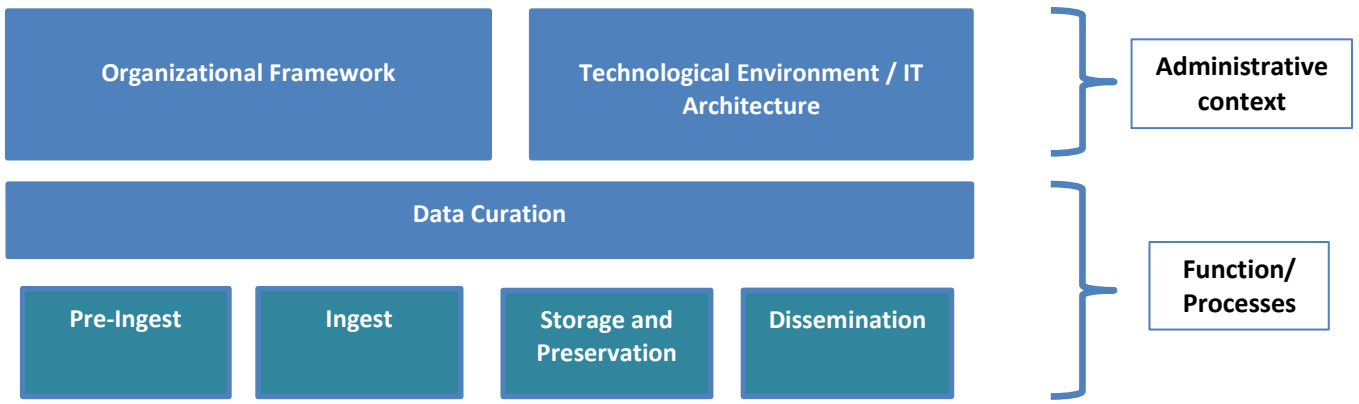
1.5 Methodology and procedure

For this report we first conducted an inventory of the commonly accepted sets of guidelines and framework models within the data preservation and data management community. The focus was restricted to reports or models which were purposed to be generally applicable for different research fields and different data preservation/data archiving settings. Practical reports of single organizations, such as best practices, were not included in this analysis (see D4.2 for more details on single organizations and (best) practices).

Selection of best practices/guidelines was carried out through a 'purposeful', two-step sampling procedure. Firstly, we provided a broad list based on 'brainstorm' suggestions from the task participants. Secondly, we carefully filter out the most relevant through a quality assessment procedure and guidelines and framework models that have reached a certain level of maturity were included and assessed based on their differences and commonalities. Then the selected key guidelines and framework models were distributed amongst partners for further assessment and detailed analysis.

Based on input from all involved partners we were able to identify and group together a general set of attributes and activities in the data repository infrastructure. On the highest administrative or organizational level, we identified a set of managing and key supporting processes areas, including strategies, policies, finances, etc. Secondly, we identified a set of attributes concerning the technological environment. Thirdly, on the level below the administrative and organizational context, we identified a set of processes, or key practices and procedures involved in the data repository data curation activities. These data repository attributes can be summarized in a simple model.

Figure 1: Basic Data Repository Attributes



After these initial processes we discussed each of the described guidelines found in the collected material and grouped together the different descriptions and recommendations that seemed to refer to the same repository attributes. The collected material was compiled in to a document listing all the recommendations from the assessed sources. The document is available as a separate appendix (reference/link)[1].

Based on the collected material we tried to organize the material into *trust maturity levels* that are intended to work as a step-wise implementation tool for the data preservation community. The level of abstraction is high; the processes described are relevant for most information processing entities, not just data repositories within the SSH community.

As such, this report does not present a new set of guidelines, or reference framework. Rather, we try to synthesize the essence of each of the carefully selected framework and provide a 'linear' step-by-step implementation tool, for existing and emerging data repositories.

The trust maturity levels are based on the five organizational stages of digital preservation (Kenney & McGovern, 2003), the CMMI five organizational maturity levels (SEI/Carnegie Mellon, 2010), and the emerging Trusted Digital Repository (TDR, 2012) framework that will consist of three levels of increasing trustworthiness. Based on the guidelines and recommendation found in our analyses of reference frameworks we ended up with five levels in our trust maturity model. Level 1 is *OAIS Core Conformance*; level 2 is *Initial self-assessment with PLATTER/DRAMBORA*; level 3 is *Peer-reviewed self-assessment with DSA*; level 4 is *Peer-reviewed self-assessment with ISO 16363/DIN 31644*; and level 5 is *Optimization and Formal Certification - Full Conformance to ISO 16363/DIN 31644*.

Table 1: Tentative mapping of process models

<table>
<tr><th>Five-step Trust Maturity Model</th><th>Kenney & McGovern</th><th>CMMI</th><th>TDR</th></tr>
<tr><td rowspan="3">1. OAIS Core Conformance</td><td>1. Acknowledge</td><td rowspan="3">1. Initial</td><td rowspan="3">(Assumes some organizational maturity)</td></tr>
<tr><td>2. Act</td></tr>
<tr><td>3. Consolidate</td></tr>
<tr><td>2. Initial self-assessment, PLATTER/DRAMBORA</td><td rowspan="2">4. Institutionalize</td><td>2. Managed</td></tr>
<tr><td>3. Peer-reviewed self-assessment I, DSA</td><td>3. Defined</td><td>1. Basic Certification</td></tr>
<tr><td>4. Peer-reviewed self-assessment II, ISO 16363/DIN 31644</td><td rowspan="2">5. Externalize</td><td>4. Quantitatively Managed</td><td>2. Extended Certification</td></tr>
<tr><td>5. Optimization and Formal Certification</td><td>5. Optimizing</td><td>3. Formal Certification</td></tr>
</table>

[1] Note that in addition to the guidelines and frameworks presented in this chapter, other sources were assessed in an early phase of the compiling of the report. These include the final report from the Blue Ribbon Task Force on Sustainable Digital Preservation (DigiPlanet); the RLG-OCLC report on attributes and responsibilities of trusted digital repositories; the Digital Preservation Handbook from the Digital Preservation Coalition; and the Curation Lifecycle Model from the Digital Curation Centre. These guidelines have a significant overlap with other guidelines, or lack a clear audit/certification perspective. More information about all the sources assess are included in the appendix to this report.

2. Guidelines and Frameworks

2.1 Key Guidelines and Frameworks

The following segment presents a brief overview and description of the reference frameworks and guidelines that constitutes the five-level trust maturity model.

2.1.1 OAIS - Reference Model for an Open Archival Information System (OAIS)

The Open Archival Information System Reference Model (CCSDS 650.0-M-2, 2012) provides a high-level reference model or framework identifying the participants in digital preservation, their roles and responsibilities, and the kinds of information to be exchanged during the course of deposit and ingest into and dissemination from a digital repository.

OAIS is a recommendation of practice for providing long term preservation of digital information. An organization can establish an OAIS archive by adhering to the recommendations and standards stated by OAIS.

OAIS is currently the benchmark standard for the construction of data archiving preservation environments. This is acknowledged in projects like for example CASPAR[2], PARSE-Insight[3], DRIVER[4], SHAMAN[5], ERPANET[6], PLANETS[7] and APARSEN[8]. It is also the baseline for several of the audit and assessment tools discussed in this report, like the PLATTER, DRAMBORA, DSA and ISO 16363. Hence, the OIAS model should not be considered as a guideline as such, but more as a platform of the full 5-level maturity model.

An earlier version of the OAIS Model report contained a roadmap which included the need for a certification standard. The initial work was to be carried out outside CCSDS and then brought back into CCSDS to develop it further into an ISO standard. In 2003, Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) created a joint task force to specifically address digital repository certification. That task force published Trustworthy Repositories Audit & Certification: Criteria and Checklist (CRL, 2007), on which the ISO 16363 Recommended Practice is based (see 2.1.7 for more information on ISO 16363).

2.1.2 PLATTER - Planning Tool for Trusted Electronic Repositories

The PLATTER is a framework which provides a basis for a digital repository to plan the development of its goals, objectives and performance targets over the course of its lifetime in a manner which will contribute to the repository establishing trusted status amongst its

[2] Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval. http://www.casparpreserves.eu/index.html
[3] Permanent Access to the Records of Science in Europe. http://www.parse-insight.eu/
[4] Digital Repository Infrastructure Vision for European Research. http://www.driver-community.eu/
[5] Sustaining Heritage Access through Multivalent Archiving. http://shaman-ip.eu/
[6] Electronic Resource Preservation and Access Network. http://www.erpanet.org/index.php
[7] Preservation and Long-term Access through Networked Services. http://www.planets-project.eu/
[8] Alliance for Permanent Access to the Records of Science in Europe. http://www.alliancepermanentaccess.org/index.php/aparsen/

stakeholders. PLATTER should not be perceived as an audit or certification tool itself, but rather a framework that will allow new repositories to incorporate the goal of achieving trust into their planning from an early stage. A repository planned using PLATTER will find itself in a better position when it subsequently comes to apply one of the existing auditing tools to confirm the adequacy of its procedures for maintaining the long term usability of and access to its material.

The PLATTER acknowledges the diversity of the organizations which may be included under the term "digital repository" by requiring repositories to answer a questionnaire which characterizes the repository relative to other repositories, and which can be used to determine how and whether the identified goals and objectives are to be realized in a given organization.

The PLATTER process is centered on a group of Strategic Objective Plans (SOPs) through which a repository specifies its current objectives, targets, or key performance indicators in those areas which have been identified as central to the process of establishing trust.

2.1.3 DRAMBORA - Digital Repository Audit Method Based on Risk Assessment

Developed jointly by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE), the DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) toolkit represents the outcome of a period of pilot repository audits undertaken by the DCC throughout 2006 and 2007. It presents a methodology for self-assessment, encouraging organizations to establish a comprehensive self-awareness of their objectives, activities and assets before identifying, assessing and managing the risks implicit within their organization. DRAMBORA focuses on self-assessment and risk-management instead of setting rules for data ingestion.

The toolkit is intended to facilitate internal audit by providing repository administrators with a means to assess their repository's capabilities, identify their weaknesses, and recognize their strengths. The toolkit is especially suitable for digital repositories that are still in their infancy or which has dispersed its activities into disorganized entities. As such, this model is intended to be responsive to the rapidly developing data repository landscape.

2.1.4 DSA - Data Seal of Approval

Based on NESTOR, DRAMBORA, TRAC, and other initiatives like the Foundations of Modern Language Resource Archives of the Max Planck Institute (Wittenburg, Broeder, Klein, Levinson, & Romary) and Stewardship of Digital Research Data: A Framework of Principles and Guidelines[9] published by the Research Information Network, DANS prepared a 'distilled' minimum set of criteria for digital research data in 2008. These were further developed by a number of institutions committed to the long-term archiving of data, and in 2009 the Data Seal of Approval was handed over to an international Board.

By granting the Seal, the DSA initiative seeks to support the durability of the data concerned, and to promote the goal of durable archiving in general.

[9] http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines.

The Data Seal of Approval is granted to repositories maintained by organizations that are committed to archiving and providing access to data for the long term, as evidenced by peer-reviewed self-assessments, which are processed through the DSA online assessment tool.

DSA is a set of 16 guidelines for data repositories or data archives. It regards three stakeholders of data: the data producer, the data consumer and the data archive. Hence the baseline of the guidelines/criteria maps closely to the core functions in the OAIS reference model.

As part of the construction of CESSDA ERIC, CESSDA member institutions will conduct a conformity assessment project starting February 2013.The list of repositories that have in the past acquired the Data Seal of Approval includes DANS Electronic Archiving System (EASY), German National Library (DNB), Inter-university Consortium for Political and Social Research (ICPSR), The Language Archive (Max Planck Institute for Psycholinguistics) and UK Data Archive, amongst others.

2.1.5 NESTOR - Catalogue of Criteria for Trusted Digital Repositories

The aim of the NESTOR catalogue is to "…formulate criteria which can be used for a broad spectrum of digital long-term repositories and which will seek to retain their validity over a longer period. The assumption is therefore that relatively abstract criteria will need to be chosen" (NESTOR, 2009).

The criteria are defined in close collaboration with a wide range of different memory organizations, information producers, experts and other interested parties. This open approach is the basis for achieving a high degree of universal validity and practical applicability and facilitates broad-based acceptance of the results of any evaluations conducted on the basis of these criteria (CESSDA PPP, 2010).

The NESTOR catalogue uses 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 data access. Hence it applies the information package definitions and terminology from the OAIS model (submission information package - SIP for ingest; archival information package - AIP for archival storage; and dissemination information package - DIP for access).

2.1.6 DIN 31644: Information and documentation - Criteria for trustworthy digital archives.

The DIN working group on Trustworthy Digital Archives developed the DIN standard 31644 which is a set of criteria that define standardized requirements for the setup and management of digital archives.

The DIN working group on Trustworthy Digital Archives developed the DIN standard 31644 based on 10 core requirements developed by this international collaboration. The scope was broadened from institutions such as archives, libraries and museums to all institutions who aim to preserve information in digital form. The main part of the standard consists of 34 requirements structured in 3 parts: organisation, management of intellectual entities and their representations, and infrastructure and security. Appendices with examples of digital archives and best practices for each requirement as well as literature complete the standard. The DIN standard was published in its final form in German in 2012 after a final test process undertaken as part of the APARSEN project and plans for a translation are ongoing (UKDA, 2012).

2.1.7 ISO 16363: Audit and Certification of Trustworthy Digital Repositories

The Audit and Certification of Trustworthy Digital Repositories (CCSDS 652.0-M-1, 2011) document is a recommendation to be used as the basis for providing audit and certification of the trustworthiness of digital repositories. It provides a detailed specification of criteria by which digital repositories can be audited.

The document is designed to be used as a tool for audit and certification processes for digital repositories which seek to assess and demonstrate their trustworthiness. Building on the RLG-OCLC Report “Trusted Digital Repositories: Attributes and Responsibilities” (RLG-OCLC, 2002), and later on the “Trustworthy Repositories Audit & Certification: Criteria and Checklist” (CRL, 2007), it presents high-level criteria to assess the organizational infrastructure, digital object management, and infrastructures and security risk management in digital repositories. As of 2012 it has been brought back into CCSDS and has become an official ISO standard (ISO 16363:2012). On connections to the OAIS model, see above.

3. Compiling the Frameworks: A Five-Level Trust Maturity Development Model

In this segment we present the five steps of the trust maturity model in more detail, providing descriptions of guidelines and pointers to supporting tools and sources within each level. The level of abstraction is generally high; the processes described are relevant for most information processing entities, not just data repositories within the SSH community.

The *key indicators* segment is meant as example descriptions of repository status, mostly drawn from the CMM-I and Kenny & McGovern models. In addition each level contains an overview of the *key guidelines,* which provide more detailed descriptions of the recommended guidelines involved in the specific level. It also provides pointers to the *sources of key guidelines* and a list of *supporting guidelines and standards*.

3.1 Trust Maturity Level 1: OAIS Core Conformance

3.1.1 Key Indicators

At trust maturity level 1 purpose, scope, objectives and goals of the repository may exist implicit, or may not be stated at all. Policies are often non-existent or may be implicit. When they do exist, policies tend to be high-level, as in the organization acknowledges the need to address digital preservation. The technological infrastructure may be non-existent or, if it exists, is likely to be heterogeneous, disparate and decentralized (as opposed to distributed).

Processes may only be partially performed, and/or are ad hoc and 'chaotic'. The repository usually does not provide a stable environment to support processes. Data are managed intuitively at project level without clear routines and practices. Repositories are capable of producing services that work, but they may frequently exceed their budget and schedules. Designated communities and significant properties are partially or implicitly defined, making the organizational processing focus heavily reactive, rather than proactive.

Repositories at this maturity level are not necessarily emerging or recently established organizations; they may also be existing experienced repositories that have dispersed into several heterogeneous tasks and processes that have led the organization to loose overview of its core activities and basic responsibilities, and that wants to 'reestablish' itself.

At trust maturity level 1 the repository should recognize that there is a problem and acknowledging the need to act, in the self-interest of the repository. The repository should

recognize the content of the core elements of their activity, that is identifying the broader organizational context. The OAIS core elements may work as a proper guideline to framing the core activities of the organization and recognizing its commitments and responsibilities.

3.1.2 Key Guidelines

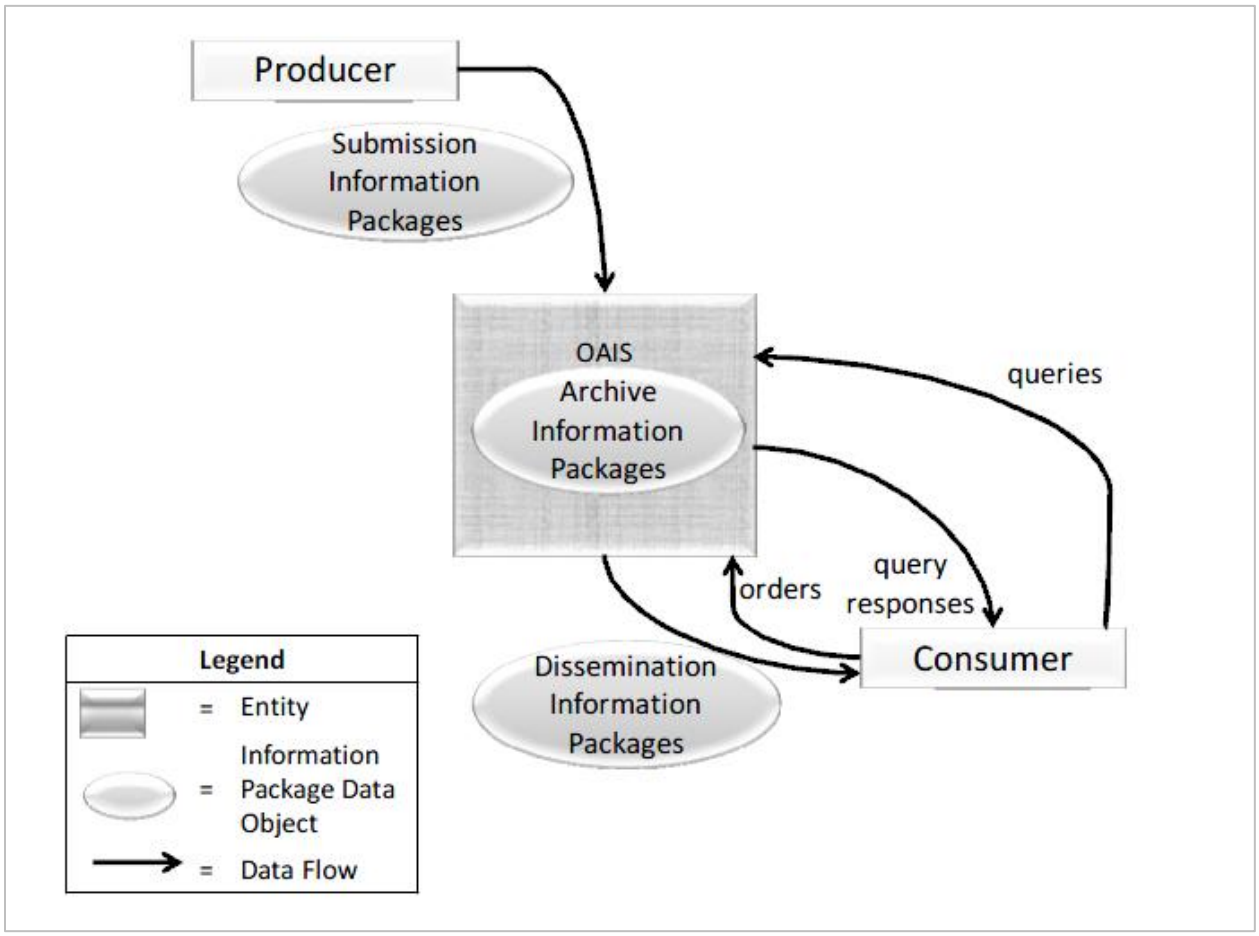
OAIS itself defines what conformance involves:

- A conforming OAIS Archive implementation shall support the model of information described in 2.2. (OAIS Information Definition section). The OAIS framework recognizes a clear definition of information as central to the ability of an OAIS to preserve it.

- A conforming OAIS Archive shall fulfill the responsibilities listed in 3.1 (Mandatory Responsibilities).

At the core of the OAIS model is recognition of the information processes involved in the data preservation. That is, accepting information from appropriate producers (ingest), controlling, managing and preserving the received information (archiving), and make this information available (disseminate) to relevant users, the designated community. As such, to support the OAIS Information Model it is necessary to distinguish between an *Information Package* that is preserved by an OAIS and the Information Packages that are submitted to, or disseminated from, an OAIS. The *Submission Information Package (SIP)* is that package that is sent to an OAIS by a Producer. Its form and detailed content are typically negotiated between the Producer and the OAIS. Within the OAIS one or more SIPs are transformed into one or more *Archival Information Packages (AIPs)* for preservation.

Figure 1: The simple OAIS model of information



It is important to notice that the OAIS Reference Model does not define or require any particular method of implementation of these concepts. As such OAIS takes a very general definition of its prime concerns; the terminology introduced is designed to be widely applicable.

In addition to supporting the Information Model, a conforming OAIS Archive shall fulfill the 'mandatory' responsibilities listed below. In order to operate an OAIS Archive, an organization must comply with these responsibilities:

- Negotiate for and accept appropriate information from information Producers.
- Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.

- Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.

- Ensure that the information to be preserved is Independently Understandable to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information.

- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed as part of an approved strategy. There should be no ad-hoc deletions.

- Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity.

By supporting the OAIS information model and acknowledging the OAIS Archive responsibilities, the repository services will make capturing, storing, maintaining, and providing access to digital resources an integral part of the organization.

Supporting and acknowledging does not mean 'conformance' in its strict form (compatibility of observations) but is to be understood more as a 'correspondence in form, manner, or character'. Full formal conformance should be achieved only through external audit or certification. Hence, here we consider informal conformance as a first step towards formal conformance.

Digital archives sometimes claim to be conformant with OAIS when in fact what they mean is that they can use OAIS terminology to describe their functions. This is not actually conformance; it just means that OAIS terminology might be very useful (Giaretta, 2011).

3.1.3 Key Guideline Sources

- Support and acknowledge OAIS Information Model: *Section 2.2 of CCSDS 650.0-M-2*
- Conforming OAIS Archive Responsibilities: *Section 3.1 of CCSDS 650.0-M-2*

3.1.4 Supporting Guidelines and Standards

There are numerous sources that can provide additional guidance at the initial phase of trust development. We have listed a selection here, all of which have a general scope and/or a data life-cycle approach. Although some of them have are designed for statistical organizations (e.g.

GSBPM, CVD) the processes describes in these sources can be considered more generally and be of value for all kinds of information processing systems that supports and acknowledges an OAIS information model platform.

- Preservation Management of Digital Materials: The Handbook[10].

  The handbook is intended to provide a bridge between broad, high level overviews and explicit, detailed guidelines applicable to the needs of a specific institution. The strategic overviews are intended to link to operational activities in order to reinforce the need to develop practical procedures grounded firmly in the business mission of the institution.

- Trustworthy Information Systems Handbook. Version 4. Saint Paul, Minnesota: Minnesota Historical Society, July 2002[11].

  Intended for government agencies; provides a practical set of tools to craft procedures based on the specific and unique needs and information requirements of information systems.

- DCC Curation Lifecycle Model[12]

  The model provides an overview of the stages involved in curation and preservation of data, from initial conceptualization or receipt through the iterative curation cycle. The model can be used to plan activities within information organizations to ensure that all of the necessary steps in the curation lifecycle are covered.

- Generic Statistical Business Process Model (GSBPM).[13]
- The Eurostat “Cycle de Vie des Données" (CVD) model.
- DDI 3.0 Combined Life Cycle Model[14].
- Statistical Data and Metadata eXchange (SDMX).[15]

[10] http://www.dpconline.org/advice/preservationhandbook
[11] http://www.mnhs.org/preserve/records/tis/tis.html
[12] http://www.dcc.ac.uk/resources/curation-lifecycle-model
[13] http://www.unece.org/stats/gsbpm
[14] http://www.ddialliance.org/Specification/DDI-Lifecycle/
[15] The SDMX standards do not provide a model for statistical business processes in the same sense as GSBPM, CVD and DDI. However they do provide standard terminology for statistical data and metadata, as well as technical standards and content-oriented guidelines for data and metadata transfer, which can also be applied between sub-processes within a statistical organization. http://sdmx.org/

3.2 Trust Maturity Level 2: Initial self-assessment, PLATTER/DRAMBORA

3.2.1 Key Indicators

At maturity level 2 the repository makes explicit its commitment to digital preservation by developing basic, essential policies and by understanding the value of policies as part of the solution. The repository has ensured that processes are planned and executed in accordance with policy, and the projects are assigned with adequate resources to produce controlled outputs. Although the activities may still be heterogeneous and dispersed among various projects, there is a broadening of the digital preservation scope from the project level towards a homogenization and institutionalization of projects and processes. Commitments are established among relevant stakeholders and are revised as needed, through monitoring, controlling and reviewing. Processes, work products and services satisfy their specified process descriptions, standards, and procedures.

3.2.2 Key Guidelines

- Self-assessment through PLATTER and DRAMBORA

The discussion of the Strategic Objective Plans (SOPs) in PLATTER lists a large number of points to be addressed in the development of detailed objectives and targets. These are extensively drawn from the TRAC (CRL, 2007) and NESTOR (NESTOR, 2009) checklists, though it is not identified which concepts have been drawn from TRAC and which from NESTOR or other sources.

However, a repository seeking to satisfy the ISO 16363/DIN 31644 or NESTOR checklists comprehensively can use PLATTER as a preparation or during the planning stage, as PLATTER try to be comprehensive, in the sense that it is covering all the major points identified in the mentioned checklists. Thus any repository which adopts the PLATTER framework (that is, all points covered in the discussion of the SOPs must be addressed), will have covered the majority of the points made in both the ISO 16363 and NESTOR checklists.

Before moving on to Trust Maturity Level 3 and self-assessment to the ISO 16363, firstly, at this level, the repository can use PLATTER in combination with the DRAMBORA self-audit tool as an initial step towards process improvement. PLATTER is designed to complement DRAMBORA and a repository planned using PLATTER will be prepared to use the DRAMBORA as a self-assessment tool (DCC & DPE, 2007). In DRAMBORA, the initial stages of the risk analysis require the repository to identify and document its goals.

Through DRAMBORA the repository proceeds to describe the activities it undertakes in pursuit of those goals and the assets which it deploys. The DRAMBORA risk analysis then consists of identifying threats to the achievement of those goals. When PLATTER has been used for

planning of objectives, a repository will be in a position to carry out a DRAMBORA analysis because all its objectives should be documented. The combination of PLATTER and DRAMBORA may therefore represent a useful tool in the initial steps of trust development.

3.2.3 Key Guideline Sources

- PLATTER Self-assessment questions (see appendix 1). (DPE, 2008).
- DRAMBORA Self-assessment questions (see appendix 2). (DCC & DPE, 2007).

3.2.4 Supporting Guidelines and Standards

- "Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems." Digital Preservation Management Resources. Cornell University[16].

  > Online tutorial developed for the Digital Preservation Management workshop, developed and maintained by Cornell University Library, 2003-2006; extended and maintained by ICPSR, 2007-2012; and now extended and maintained by MIT Libraries, 2012-on.

- Green, McDonald & Rice, 2009. "Policy-making for Research Data in Repositories. A Guide". Version 1.2. Data Information Specialists Committee –UK[17].

  > Intended to be used as a decision-making and planning tool for institutions with digital repositories in existence or in development that are considering adding research data to their digital collections.

- Chang, Debbie (2010). "TAPS: Checklist for Responsible Archiving of Digital Language Resources"[18].

  > TAPS is a master thesis by Debbie Chang from the Faculty of the Graduate Institute of Applied Linguistics in Texas, USA. It develops a checklist that is intended to help depositors of language materials to assess digital language archives based on areas of special concern to linguists and language communities and to recommend best practices for the long-term preservation of digital information. The TAPS Checklist

[16] http://www.dpworkshop.org/index.html
[17] http://www.disc-uk.org/docs/guide.pdf
[18] http://www.gial.edu/images/theses/Chang_Debbie-thesis.pdf

was formulated by the author through a comparison of components of TRAC (CRL, 2007), and earlier versions of NESTOR (NESTOR, 2009) and the DSA (DSA, 2010).

3.3 Trust Maturity Level 3: Peer-reviewed self-assessment, DSA

3.3.1 Key Indicators

At maturity level 3, processes are well characterized and understood, and are described in standards, procedures, tools, and methods. These standard processes are used to establish consistency across the organization. Projects establish their defined processes by tailoring the organization's set of standard processes according to guidelines. Organization-wide entities that coordinate, authorize, and mandate digital preservation programs may be established, or some equivalent mechanism that allows for consistent and systematic management rather than event-based, reactive responses; planning and management are characterized by *responding to* rather than *reacting to*, and are *proactively* anticipating needs. The organization has made itself and its services subject to relevant self-monitoring and measuring (PLATTER, DRAMBORA), and expectations that these services will be reliable and consistent has become evident.

3.3.2 Key Guidelines

At this stage the archive or repository can aim for the DSA. Achieving the DSA means that the data archive or repository is in compliance with the sixteen DSA guidelines as determined through an assessment procedure. Although these guidelines pertain to three stakeholders – the data producer (three guidelines), the data consumer (three guidelines) and the data archive (ten guidelines) – the data archive is seen as the primary implementer of the guidelines. The data archive as an organization should assume responsibility for the overall implementation of the DSA in its own specific field (DSA, 2010).

The completion of the DSA self-assessment form is the starting point for the reviewing procedure. The assessment is then sent to the Board in order to decide via peer review whether an organization will be granted the Seal of Approval. There is no audit, no certification: just a review on the basis of trust. The assessment lists the sixteen Data Seal of Approval guidelines. In the assessment, the organization describes how these guidelines relate to the repository and how they have been implemented (DSA, 2010).

It is important to note that this is not an audit, or a certification, just a review based on trust. The peer review is performed by a member of the DSA Board or another qualified individual appointed by the Board, depending on the subject matter coverage of the repository.

3.3.3 Key Guideline Sources

- The sixteen Data Seal of Approval guidelines (see appendix 3). (DSA, 2010).

3.3.4 Supporting Guidelines and Standards

- NESTOR criteria: Catalogue of Criteria for Trusted Digital Repositories. Version 2. Published by NESTOR Working Group Trusted Repositories – Certification. Frankfurt am Main, November 2009.

3.4.4 Trust Maturity Level 4: Peer-reviewed self-assessment, ISO 16363/DIN 31644

3.4.1 Key Indicators

At maturity level 4, processes are well characterized and understood, and are described in standards, procedures, tools, and methods. These standard processes are used to establish consistency across the organization. Projects establish their defined processes by tailoring the organization's set of standard processes according to guidelines. Organization-wide entities that coordinate, authorize, and mandate digital preservation programs may be established, or some equivalent mechanism that allows for consistent and systematic management rather than event-based, reactive responses; planning and management are characterized by *responding to* rather than *reacting to*, and are *proactively* anticipating needs. The organization has made itself and its services subject to relevant self-monitoring and measuring (PLATTER, DRAMBORA) and/or peer-reviewed self-assessment through DSA, and expectations that these services will be reliable and consistent has become evident.

3.4.2 Key Guidelines

- Conformance to the full OAIS Model (detailed functional entities)
- Self-audit with the ISO 16363 / DIN 31644

In general, the ISO 16363 is closely mapped to the OAIS Reference Model; key terms in the document have been adopted from the OAIS, which has become a foundational document for digital preservation (see above). Conformance to all functional entities in the OAIS Model is not to be taken as a recommended design or implementation, and actual implementations are not expected to have a one-to-one mapping to all functions; a repository may for example choose to combine functions or break out functionalities differently (CCSDS 650.0-M-2, 2012). However, it is recommended at this level to map the repository activities to the full functional OAIS model, as there is a close relationship between the OAIS functional entities and the ISO

16363 checklist. Mapping activities to the OAIS functional entities can be applied as a preparation to, or simultaneously with, the self-auditing to the ISO 16363/DIN 31644 checklists.

Conformance to the full OAIS Model involves mapping all repository activities to the detailed six functional entities and related interfaces. These are: Ingest Functional Entity; Archival Storage Functional Entity; Data Management Functional Entity; Administration Functional Entity; Preservation Planning Functional Entity; and Access Functional Entity. Sub-processes and specific flows of information among the functional entities are identified in the full functional entities model. However, the OAIS report states that the functional entities are not to be taken as a recommended design or implementation, and actual implementations are not expected to have a one-to-one mapping to the functions described. Data repositories may for example choose to combine functions or break out functionality differently (CCSDS 650.0-M-2, 2012).

3.4.3 Key Guideline Sources

- CCSDS 650.0-M-2, Reference Model for an Open Archival information System (Magenta Book, June 2012). Also available as ISO 14721:2012.

- CCSDS 652.0-M-1, Audit and Certification of Trustworthy Digital Repositories (Magenta Book, Issue 1). Also available as ISO 16363:2012.

- DIN 31644: Information and documentation - Criteria for trustworthy digital archives.

3.4.4 Supporting Guidelines and Standards

- NESTOR criteria: Catalogue of Criteria for Trusted Digital Repositories. Version 2. Published by NESTOR Working Group Trusted Repositories – Certification. Frankfurt am Main, November 2009.

3.5 Trust Maturity Level 5: Optimization and Formal Certification - Full Conformance to ISO 16363/DIN 31644

3.5.1 Key Indicators

At maturity level 5, a repository continually seeks to improve its processes based on a broad (quantitative) understanding of its business objectives and performance needs. The repository may use a quantitative approach to understand the variation inherent in the process and the causes of process outcomes.

For selected sub-processes, the repository may seek specific process improvement. When selecting sub-processes improvement, it is critical to understand the relationships between different sub-processes and their impact on achieving the objectives for quality and process performance.

Maturity level 5 focuses on continually improving process performance through process and technological improvements. The repository's quality and processes are continually revised to reflect changing business objectives, repository performance and technological changes.

Simultaneously with process improvement the repository may seek to obtain full external audit and certification based on ISO 16363 or the equivalent DIN 31644.

3.5.2 Key Guidelines

- Full conformance to ISO 16363 / DIN 31644.

An archive that wants to reach full conformance to the ISO 16363 / DIN 31644 must have satisfied an external auditor on each of the requirements. However, the aim of the audit process is to create a process of continuous improvement. Thus the outcome of the audit will not be a simple yes/no but rather a judgment about areas that need improvement. Attaining trustworthy status is not a one-time accomplishment; to retain trustworthy status, a repository will need to undertake a regular cycle of audit and/or certification. An important part of the cycle of auditing (either self-auditing or through external assessments) is through more detailed auditing of the various sub-processes of the repository activities.

3.5.3 Key Guideline Sources

- CCSDS 650.0-M-2, Reference Model for an Open Archival information System (Magenta Book, June 2012). Also available as ISO 14721:2012.

- CCSDS 652.0-M-1, Audit and Certification of Trustworthy Digital Repositories (Magenta Book, Issue 1, September 2011). Also available as ISO 16363:2012.

- DIN 31644: Information and documentation - Criteria for trustworthy digital archives.

3.5.4 Supporting Guidelines and Standards

- Quality Management Systems—Fundamentals and Vocabulary. International Standard, ISO 9000:2005. 3rd edition. Geneva: ISO, 2005.

  ISO 9000 addresses quality assurance components within an organization and system management. However, it is not specifically developed to gauge the trustworthiness of organizations operating digital repositories.

- Information Technology—Security Techniques—Code of Practice for Information Security Management. International Standard, ISO/IEC 17799:2005. 2nd edition. Geneva: ISO, 2005.

  ISO 17799:2005 was developed specifically to address data security and information management systems. Like ISO 9000, it was not designed to address the trustworthiness of digital repositories. Its requirements for information security seek data security compliance to a very granular level, but do not address organizational, procedural, and preservation planning components necessary for the long-term management of digital resources.

- Information and Documentation—Records Management—Part 1: General. International Standard, ISO 15489-1:2001. Geneva: ISO, 2001.

- Information and Documentation—Records Management—Part 2: Guidelines. International Standard, ISO/TR 15489-2:2001. Geneva: ISO, 2001.

  ISO 15489-1:2001 and ISO 15489-2:2001 define a systematic and process-driven approach that governs the practice of records managers and any person who creates or uses records during their business activities, treats information contained in records as a valuable resource and business asset, and protects/preserves records as evidence of actions. Conformance to ISO 15489 requires an organization to establish, document, maintain, and promulgate policies, procedures, and practices for records management, but, by design, addresses records management specifically rather than applying to all types of repositories and archives.

- ISO 20652:2006 Space data and information transfer systems—Producer-Archive Interface—Methodology Abstract Standard (PAIMAS) (the more specific Producer-Archive Interface Specification (PAIS) is under preparation).

  Identifies, defines and provides structure to the relationships and interactions between an information producer and an archive. It defines the methodology for the structure of actions that are required from the initial time of contact between the producer and the archive until the objects of information are received and validated by the archive.

- NESTOR-studies 10 (2009): Into the Archive - A guide for the information transfer to a digital repository.

  Into the Archive" aims to clarify the goals and unique aspects of ingesting information into a digital repository. The guide draws heavily on OAIS and PAIMAS.

- PREMIS Data Dictionary for Preservation Metadata. Version 2.0, PREMIS Editorial Committee, March 2008 [19].

  The PREMIS Data Dictionary is a comprehensive, practical resource for implementing preservation metadata in digital archiving systems. The Dictionary builds on the OAIS reference model.

- Data Documentation initiative (DDI) [20].

  DDI is an international standard for describing data from the social, behavioral, and economic sciences. Expressed in XML, the DDI metadata specification version 3 supports the research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving.

- The Dublin Core Metadata Initiative (DCMI)[21].

  The Dublin Core Metadata Initiative provides core metadata vocabularies in support of interoperable solutions for discovering and managing resources.

[19] http://www.loc.gov/standards/premis/v2/premis-2-0.pdf
[20] http://www.ddialliance.org/
[21] http://dublincore.org/

- Statistical Data and Metadata Exchange (SDMX).
- DIN 31646:2012 Information and documentation - Requirements for the long-term management of persistent identifiers.
- ISO 30301:2011 Information and documentation — Management systems for records — Requirements - First Edition.
- ISO/IEC 27001:2005 - Information technology - Security techniques - Specification for an information security management system.
- ISO/IEC 27002:2005 - Information technology - Security techniques - Code of practice for information security management.
- NESTOR (2009): Catalogue of criteria for assessing the trustworthiness of PI systems. Draft for public comment. NESTOR-studies 13[22].

> Catalogue of criteria for assessing the trustworthiness of persistent identifier systems (PI systems). The catalogue is intended to help providers and users of persistent identifiers) to keep digital objects identifiable, referenceable and accessible over longer periods and despite unforeseeable changes.

[22] http://nbn-resolving.de/urn:nbn:de:0008-20080710227

4. Summary

In the Social Sciences and Humanities research is increasingly driven by the availability of a variety of digital resources, which exhibit an escalating internal complexity as well as multifarious external relationships. The data production, management and dissemination processes are organized in a distributed manner, both within and between repositories, which may lead to a situation of fragmentation that should be taken into account when designing or developing research infrastructures and data repositories for the corresponding scientific disciplines. The data produced by the different data communities should be made persistently available to their respective designated communities via environments that implement discipline specific workflows and traditions in a trustworthy manner.

This report presents the results of the assessment of existing data repository model frameworks for self-assessment and full external certification of data repositories. The assumption is that all data repositories that seek to become trusted digital repositories should assess their organization and aim for a certain level of trust maturity.

Formal certification may not be relevant for all repositories but we believe that the five level Trust Maturity Model contains valuable guidelines for the continued viability and trust of digital materials and may serve as a baseline and a roadmap for improving the state of long term data preservation in the SSH domain.

Table 1: Summary of maturity levels and key guidelines

Trust Maturity Level	Key Guideline	Guideline Source
1. OAIS Core Conformance	Support OAIS Information Model. Acknowledge OAIS Archive responsibilities.	OAIS Information Model: Section 2.2 of CCSDS 650.0-M-2 / ISO 14721:2012. OAIS Archive Responsibilities: Section 3.1 of CCSDS 650.0-M-2 / ISO 14721:2012.
2. Initial self-assessment, PLATTER/DRAMBORA	Self-assessment through PLATTER and DRAMBORA.	PLATTER Key Self-assessment questions. DRAMBORA Key Self-assessment questions.
3. Peer-reviewed self-assessment I, DSA	Peer-reviewed self-assessment I, DSA.	Data Seal of Approval Guidelines. Support: NESTOR criteria
4. Peer-reviewed self-assessment II, ISO 16363/DIN 31644	Conformance to the OAIS Detailed Functional Model. Self-audit with the ISO 16363. Alternatively, self-audit with DIN 31644.	OAIS Detailed Functional Model: Section 4.1 of CCSDS 650.0-M-2 / ISO 14721:2012. CCSDS 652.0-M-1 / ISO 16363:2012. DIN 31644
5. Certification and Optimization	External review and formal certification in conformance with the ISO 16363. Alternatively, with DIN 31644.	CCSDS 652.0-M-1 / ISO 16363:2012. DIN 31644.

Appendix 1: PLATTER Key Self-assessment questions

REPOSITORY PURPOSE AND FUNCTION
Q1.1 Source of Mandate: What is the source of the repository's mandate?
Q1.2 Commercial Status: Is the Repository for profit or non-profit?
Q1.3 Legal acquisition rights: Does the Repository receive a significant proportion of its material from a legally mandated source (e.g. archival deposit or legal deposit)?
Q1.4 Operational Maturity: What is the operational status of the repository (not yet running, running but still under development, mature)?
SCALE OF REPOSITORY
Q2.1 Data Quantity : What is the amount of digital material you expect to archive per year (GB)?
Q2.2 Data Quantity: How many distinct digital objects do you expect to archive per year?
Q2.3 Human Size : How many fulltime-equivalent staff does the Repository expect to employ?
Q2.4 Human Size: How many distinct endusers are expected to access material in the Repository over the course of a calendar year?
OPERATION
Q3.1 Acquisition Method: Which of the three acquisition strategies (push, pull, self-creation) account for a significant portion of the total material in the Repository?
Q3.2 Data Complexity: Is the majority of the material in the Repository simple, moderately complex, or highly complex?
Q3.3 Data specialization: How specialised is the data in the Repository (low, medium or high) ?
Q3.4 Data Sensitivity: How sensitive is the most sensitive material in the Repository (low, medium, high)?
Q3.5 Access Rights: In which of the three access classes (open, restricted, closed) does the Repository have significant holdings?
TECHNICAL SOLUTIONS AND IMPLEMENTATION CHOICES
Q4.1 Source of Metadata: What are the main sources of bibliographic and descriptive metadata in the repository?
Q4.2 Interoperability Standards: What interoperability standards are implemented in the Repository?
Q4.3 Storage Strategy: What strategy is used for storage? (in-house, external, in-house under external support)
Q4.4 Software Strategy: What strategy is used for software management?

Appendix 2: DRAMBORA Key Self-assessment questions

T1: What is the mandate of your repository or the organisation in which it is embedded?
T2: List goals and objectives of your repository
T3: List your repository's strategic planning documents
T4: List the legal, regulatory and contractual frameworks or agreements to which your repository is subject
T5: List the voluntary codes to which your repository has agreed to adhere
T6: List any other documents and principles with which your repository complies
T7: Identify your repository's activities, assets and their owners
T8: Identify risks associated with activities and assets of your repository
T9: Assess the identified risks

Appendix 3: The Data Seal of Approval Guidelines

1. The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.
2. The data producer provides the research data in formats recommended by the data repository
3. The data producer provides the research data together with the metadata requested by the data repository
4. The data repository has an explicit mission in the area of digital archiving and promulgates it
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies documented processes and procedures for managing data storage
7. The data repository has a plan for long-term preservation of its digital assets
8. Archiving takes place according to explicit workflows across the data life cycle
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects
10. The data repository enables the users to utilize the research data and refer to them
11. The data repository ensures the integrity of the digital objects and the metadata
12. The data repository ensures the authenticity of the digital objects and the metadata
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS
14. The data consumer complies with access regulations set by the data repository
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and research for the exchange and proper use of knowledge and information
16. The data consumer respects the applicable licenses of the data repository regarding the use of the research data

Appendix 4: ISO 16363 Checklist

Note that the numbering is from the original document (CCSDS 652.0-M-1, 2011).

ORGANIZATIONAL INFRASTRUCTURE
GOVERNANCE AND ORGANIZATIONAL VIABILITY
3.1.1 The repository shall have a mission statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information.
3.1.2 The repository shall have a Preservation Strategic Plan that defines the approach the repository will take in the long-term support of its mission.
3.1.3 The repository shall have a Collection Policy or other document that specifies the type of information it will preserve, retain, manage, and provide access to.
3.2 ORGANIZATIONAL STRUCTURE AND STAFFING
3.2.1 The repository shall have identified and established the duties that it needs to perform and shall have appointed staff with adequate skills and experience to fulfill these duties.
3.3 PROCEDURAL ACCOUNTABILITY AND PRESERVATION POLICY FRAMEWORK
3.3.1 The repository shall have defined its Designated Community and associated knowledge base(s) and shall have these definitions appropriately accessible.
3.3.2 The repository shall have Preservation Policies in place to ensure its Preservation Strategic Plan will be met.
3.3.3 The repository shall have a documented history of the changes to its operations, procedures, software, and hardware.
3.3.4 The repository shall commit to transparency and accountability in all actions supporting the operation and management of the repository that affect the preservation of digital content over time.
3.3.5 The repository shall define, collect, track, and appropriately provide its information integrity measurements.
3.3.6 The repository shall commit to a regular schedule of self-assessment and external certification.
3.4 FINANCIAL SUSTAINABILITY
3.4.1 The repository shall have short- and long-term business planning processes in place to sustain the repository over time.
3.4.2 The repository shall have financial practices and procedures which are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.
3.4.3 The repository shall have an ongoing commitment to analyze and report on financial risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).
3.5 CONTRACTS, LICENSES, AND LIABILITIES
3.5.1 The repository shall have and maintain appropriate contracts or deposit agreements for digital materials that it manages, preserves, and/or to which it provides access.
3.5.2 The repository shall track and manage intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.
4 DIGITAL OBJECT MANAGEMENT
4.1 INGEST: ACQUISITION OF CONTENT
4.1.1 The repository shall identify the Content Information and the Information Properties that the repository will preserve.
4.1.2 The repository shall clearly specify the information that needs to be associated with specific Content Information at the time of its deposit.
4.1.3 The repository shall have adequate specifications enabling recognition and parsing of the SIPs.
4.1.4 The repository shall have mechanisms to appropriately verify the identity of the Producer of all materials.
4.1.5 The repository shall have an ingest process which verifies each SIP for completeness and correctness.
4.1.6 The repository shall obtain sufficient control over the Digital Objects to preserve them.
4.1.7 The repository shall provide the producer/depositor with appropriate responses at agreed points during the ingest processes.
4.1.8 The repository shall have contemporaneous records of actions and administration processes that are relevant to content acquisition.
4.2 INGEST: CREATION OF THE AIP
4.2.1 The repository shall have for each AIP or class of AIPs preserved by the repository an associated definition that is adequate for parsing the AIP and fit for longterm preservation needs.
4.2.2 The repository shall have a description of how AIPs are constructed from SIPs.

4.2.3 The repository shall document the final disposition of all SIPs.
4.2.4 The repository shall have and use a convention that generates persistent, unique identifiers for all AIPs.
4.2.5 The repository shall have access to necessary tools and resources to provide authoritative Representation Information for all of the digital objects it contains.
4.2.6 The repository shall have documented processes for acquiring Preservation Description Information (PDI) for its associated Content Information and acquire PDI in accordance with the documented processes.
4.2.7 The repository shall ensure that the Content Information of the AIPs is understandable for their Designated Community at the time of creation of the AIP.
4.2.8 The repository shall verify each AIP for completeness and correctness at the point it is created.
4.2.9 The repository shall provide an independent mechanism for verifying the integrity of the repository collection/content.
4.2.10 The repository shall have contemporaneous records of actions and administration processes that are relevant to AIP creation.
4.3 PRESERVATION PLANNING
4.3.1 The repository shall have documented preservation strategies relevant to its holdings.
4.3.2 The repository shall have mechanisms in place for monitoring its preservation environment.
4.3.3 The repository shall have mechanisms to change its preservation plans as a result of its monitoring activities.
4.3.4 The repository shall provide evidence of the effectiveness of its preservation activities.
4.4 AIP PRESERVATION
4.4.1 The repository shall have specifications for how the AIPs are stored down to the bit level.
4.4.2 The repository shall have contemporaneous records of actions and administration processes that are relevant to storage and preservation of the AIPs.
4.5 INFORMATION MANAGEMENT
4.5.1 The repository shall specify minimum information requirements to enable the Designated Community to discover and identify material of interest.
4.5.2 The repository shall capture or create minimum descriptive information and ensure that it is associated with the AIP.
4.5.3 The repository shall maintain bi-directional linkage between each AIP and its descriptive information.
4.6 ACCESS MANAGEMENT
4.6.1 The repository shall comply with Access Policies.
4.6.2 The repository shall follow policies and procedures that enable the dissemination of digital objects that are traceable to the originals, with evidence supporting their authenticity.
5 INFRASTRUCTURE AND SECURITY RISK MANAGEMENT
5.1 TECHNICAL INFRASTRUCTURE RISK MANAGEMENT
5.1.1 The repository shall identify and manage the risks to its preservation operations and goals associated with system infrastructure.
5.1.2 The repository shall manage the number and location of copies of all digital objects.
5.2 SECURITY RISK MANAGEMENT
5.2.1 The repository shall maintain a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant.
5.2.2 The repository shall have implemented controls to adequately address each of the defined security risks.
5.2.3 The repository staff shall have delineated roles, responsibilities, and authorizations related to implementing changes within the system.
5.2.4 The repository shall have suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an offsite copy of the recovery plan(s).

References

ACLS. (2006). *e American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences.* Hentet fra American Council of Learned Societies: http://www.acls.org/cyberinfrastructure/ourculturalcommonwealth.pdf

Anderson, W. L. (2004, December 30). Some Challenges ansd Issues in Managing, and Preserving Access to, Long-Lived Collections of Digital Scientific and Technical Data. *Data Science Journal*, ss. 191-202.

Blue Ribbon Task Force. (2010). *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information.*

CCSDS 650.0-M-2. (2012). *Reference Model for an Open Archival Information System (Magenta Book, June 2012). Also available as ISO 14721:2012.* CCSDS.

CCSDS 652.0-M-1. (2011, February). *Audit and certification of trustworthy digital repositories (Magenta Book, September 2011). Also available as ISO 16363:2012.*

CEDARS. (2012). *Cedars project (Now UK Web Archive)*. Hentet fra http://www.webarchive.org.uk/ukwa/

CESSDA PPP. (2010). *WP6 Final report: Strengthening the CESSDA RI (D6.1).* CESSDA.

CRL. (2007, February). Trustworthy Repositories Audit & Certification: Criteria and Checklist. *Version 1.0.* Chicago.

DCC & DPE. (2007, February 28). DCC and DPE Digital Repository Audit Method Based on Risk Assessment, v1.0.

DPE. (2008). *Digital Preservation Europe: D3.2 Repository Planning Checklist and Guidance.* Digital Preservation Europe.

DSA. (2010). *Data Seal of Approval*. Hentet fra http://datasealofapproval.org/

Giaretta, D. (2011). *Advanced Digital Preservation.* Berlin: Springer-Verlag Berlin Heidelberg.

Ginsparg, P. (2004, February 4). Scholarly Information Architecture, 1989-2015. *Data Science Journal, 3*, ss. 29-37.

HLEG. (2010, October). *Riding the wave. How Europe can gain fromthe rising tide ofscientific data. Final report of the High Level Expert Group on Scientific Data.* Hentet fra CORDIS: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Kenney, A. R., & McGovern, N. Y. (2003). The Five Organizational Stages of Digital Preservation. I *Digital Libraries: A Vision for the Twenty-first Century, a festschrift to honor Wendy Lougee, 2003.*

Hentet fra http://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11

King, G. (2011). Ensuring the Data-Rich Future of the Social Sciences. *Science*(331, 719).

Mois, M., Klas, C.-P., & Hemmje, M. L. (2009). Digital preservation as communication with the future. *Digital Signal Processing, 2009 16th International Conference* (ss. 1-9). Hagen: Fac. for Math. & Comput. Sci., FernUniversitat in Hagen.

NESTOR. (2009). *NESTOR criteria : Catalogue of Criteria for Trusted Digital Repositories, Version 2.* Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek. Hentet fra http://files.d-nb.de/NESTOR/materialien/NESTOR_mat_08_eng.pdf

NESTOR. (2009). *NESTOR criteria : Catalogue of Criteria for Trusted Digital Repositories, Version 2.* Frankfurt am Main: nestor c/o Deutsche Nationalbibliothek.

OECD. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding.* Hentet fra http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf

RLG-OCLC. (2002, May). Trusted Digital Repositories: Attributes and Responsibilities. *An RLG-OCLC Report*. Mountain View, CA: Research Libraries Group.

Schott, M., Dittmann, J., Vielhauer, C., Krätzer, C., & Lang, A. (2008). *Integrity and Authenticity for Digital Long-Term Preservation in IRODS Grid Infrastructure.* Hentet fra http://wwwiti.cs.uni-magdeburg.de/~vielhaue/jabreflib/[SDVK+2008].pdf

SEI/Carnegie Mellon. (2010). *CMMI (Capability Maturity Model Integration for Development), Improving processes for developing better products and services, Version 1.3.* Software Engineering Institute/Carnegie Mellon University. Hentet fra http://www.sei.cmu.edu/

Sergeant, D. (2002). *Interpretation of the OAIS Model.* Hentet fra http://www.erpanet.org/events/2002/copenhagen/presentations/dmserpanet.ppt

TDR. (2012, 12 4). *European Framework for Audit and Certification of Digital Repositories. Trusted Digital Repository - Making the digital world more reliable.* Hentet fra http://www.trusteddigitalrepository.eu/

UKDA. (2012, 12 13). *UK Data Archive: How to Curate - Standards of Trust*. Hentet fra http://data-archive.ac.uk/curate/trusted-digital-repositories/standards-of-trust?index=3

UNESCO. (2012). *The Memory of the World in the Digital age: Digitization and Preservation*. Hentet fra http://www.unesco.org/new/en/communication-and-information/events/calendar-of-events/events-websites/the-memory-of-the-world-in-the-digital-age-digitization-and-preservation/

Wittenburg, P., Broeder, D., Klein, W., Levinson, S., & Romary, L. (u.d.). Hentet fra http://arxiv.org/abs/cs/0606006