Policy on Compliance as a Trusted Digital Repository

DARIAH 2010

DARIAH
Digital Research Infrastructure for the Arts and Humanities

Table of Contents

Executive Summary

A Trusted Digital Repository (TDR) provides reliable long-term access to digital resources under its management. The reliability and trustworthiness of digital preservation programmes are very important issues to many stakeholders. 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. Trusted Digital Repositories must have the capacity of a system to operate in accordance with objectives and specifications in order to fulfill all of the requirements of maintaining the quality of the objects for which it is responsible.

The policy outlined below is intended for repositories seeking DARIAH accreditation. This policy will be a set of criteria that a repository would have to follow to comply with DARIAH requirements. The main purpose of this policy is to outline best practice guidelines that repositories seeking to be accredited by DARIAH will be expected to follow. There are seven key attributes set out in this document: Compliance with the Reference Model for an Open Archival Information System (OAIS), Administrative responsibility, Organisational viability, Financial sustainability, Technological and procedural suitability, System Security, and Procedural accountability. By ensuring all of the attributes are up to the agreed standard, repositories will be seen to be trustworthy and accredited as such. By engaging with repositories DARIAH will aim to help them provide a high-quality service to their stakeholders.

Introduction

Audience for this document

The policy outlined below is intended for repositories seeking DARIAH accreditation. The spectrum of existing digital repositories and those currently being set up is very broad and can include university libraries, research institutions, archives, museums and service providers.

Aims of policy

This policy will be a set of criteria that a repository would have to follow to comply with DARIAH requirements. New organisations which join DARIAH or begin engaging in repository type activities will be expected to meet certain minimum standards. This will enable DARIAH to vet organisations and recommend practices to be followed – all organisations should meet policy requirements in order to be counted as Trusted. Trustworthiness is not an absolute term; it depends on the goals of a particular digital repository. A trusted digital repository 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. From DARIAH's perspective, DARIAH wants to engage with repositories to help them provide a high-quality service to their stakeholders.

Scope of policy

The main purpose of this policy is to outline best practice guidelines that repositories seeking to be accredited by DARIAH will be expected to follow. This policy takes into consideration recent international approaches and findings such as the OCLC-RLG reports published by the OCLC/RLG-NARA Digital Repository Certification Task Force, the nestor criteria catalogue, DRAMBORA, Data Seal of Approval and ISO Standards. The scope of application of this document is the entire range of digital repositories.

Definition of a Trusted Digital Repository (TDR)

A digital repository is an organisation that has responsibility for the long-term maintenance of digital resources, as well as for making them available to communities agreed upon by the depositor and the repository.[1] Trustworthiness is the capacity of a system to operate in accordance with its objectives and specifications, that is, it does exactly what it claims to do. The concept trusted digital repository is used increasingly in relation to digital preservation and data curation. In 1996 the Research Libraries Group (RLG) and the Commission on Preservation and Access published Preserving Digital Information which included a clear statement about trust in digital archives:

For assuring the longevity of information, perhaps the most important role in the operation of a digital archives is managing the identity, integrity and quality of the archives itself as a trusted source of the cultural record. Users of archived information in electronic form and of

[1] An RLG-OCLC Report, *Attributes of a Trusted Digital Repository: Meeting the needs of research resources* (August, 2001) p.9 http://www.oclc.org/research/activities/past/rlg/trustedrep/attributes01.pdf accessed 4 June 2010

archival services relating to that information need to have assurance that a digital archive is what it says it is and that the information stored there is safe for the long term.[2]

The RLG-OCLC report defined a Trusted Digital Repository (hereafter TDR) as one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.[3] The Nestor working group[4] states that a trusted 'long-term digital repository is a complex and interrelated system'.[5] The TDR must accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users. It may take different forms, for example, some institutions may choose to build local repositories while others may choose to manage the logical and intellectual aspects of a repository while contracting with a third-party provider for its storage and maintenance.[6] A digital repository can declare itself trustworthy and expect potential stakeholders to accept this declaration. However, trust is built over time and a trustworthy digital repository must continue to meet specified requirements to remain trustworthy. For repositories seeking DARIAH accreditation, this will not be a single evaluation; rather it will be ongoing process between DARIAH and the repository to ensure that all the relevant policies and procedures are both of the highest quality and are up-to-date. To be considered 'trustworthy' a TDR must:

- Accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users;
- Have an organisational system that supports both the long-term viability of the repository and the digital information for which it has responsibility;
- Demonstrate that it is fiscally responsible and is sustainable;

[2] Ibid p.6

[3] An RLG-OCLC Report, *Trusted Digital Repositories: Attributes and Responsibilities* (May,2002) p.i

[4] Network of Expertise in Long-Term Storage of Digital Resources (nestor) the German competence network for digital preservation. For more information please see: http://www.langzeitarchivierung.de/eng/index.htm

[5] OCLC Report, *Trustworthy Repositories Audit and Certification: Criteria and Checklist* (February, 2007) p.3. See also the DRIVER project's publications for more information: http://www.driver-support.eu/linkspubs/driverpubs.html

[6] RLG-OCLC (2002) p.5

- Design its systems in accordance with commonly accepted conventions and standards to ensure the ongoing management, access and security of materials deposited within it;
- Establish methodologies for system evaluation that meet community expectations of trustworthiness;
- Be relied upon to carry out its long-term responsibilities to depositors and users openly and explicitly;
- Have policies, practices and performances that can be audited and measured; and
- Meet the responsibilities outlined below in section six.[7]

Attributes of a Trusted Digital Repository

The RLG-OCLC 2002 report identified seven attributes that a digital repository must have to be considered trustworthy. These attributes must accommodate all different situations and institutional responsibilities whilst providing a basis for expectations of a trusted repository.

The identified attributes are as follows:

Compliance with the Reference Model for an Open Archival Information System (OAIS)

The Open Archival Information System (OAIS) Reference Model was developed by the Consultative Committee for Space Data Systems. It is a conceptual model and reference tool for defining a digital repository. It provides a model of the environment, functions and data types required for implementing a repository. In 2002 it became an approved official ISO (International Organisation for Standardisation) Standard – 14721. More information about the OAIS Reference Model can be found here: http://public.ccsds.org/publications/archive/650x0b1.pdf. A TDR will ensure that the overall repository system conforms to the OAIS Reference Model. The OAIS Reference Model provides a common framework, which includes terminology and concepts for describing and comparing architectures and operations of digital archives.[8] In addition, the OAIS Reference Model provides both a functional model and an information model. The functional model

[7] RLG-OCLC (2002) P.5

[8] RLG-OCLC (2002) p.13

outlines the specific tasks performed by the repository such as storage access. The information model includes a model for the creation of metadata to support long-term maintenance and access.[9]

Administrative responsibility

Administrative responsibility should include a high-level statement that demonstrates a commitment to track and comply with current and emerging standards embraced by the preservation community. It should demonstrate that a TDR has a fundamental commitment to implementing the range of community-agreed standards and best practices that affect its operations. This responsibility also includes meeting appropriate national and/or international standards for the physical environment, backup and recovery procedures and security systems.10 A TDR will involve external experts in the validation and/or certification of its process and procedures. A reliable TDR will commit itself to transparency and accountability. Essentially a TDR should make an explicit commitment to the development of TDR that complies with prevailing standards, policies and practices that can be audited and measured.

Organisational viability

An organisation that wants to become a TDR must demonstrate its viability and trustworthiness. Its mission statement must reflect a commitment to long-term retention, management of, and access to digital assets on behalf of depositors and users. The organisation should attain legal status, staff, professional development and standing that will be appropriate to the range of responsibilities they are undertaking. It should establish transparent business practices and effective management policies. This attribute includes the prospect of an ongoing mandate that would support an ongoing preservation role and a demonstrated ability to put together the resources, infrastructure and work teams that could manage the complexity of digital preservation.[11]

Financial sustainability

A TDR should establish and maintain good business practices and an auditable business plan in order to demonstrate its financial sustainability over time. It should demonstrate that the organisation is able to continue to provide the required resources well into the future with a

[9] RLG-OCLC (2002) p.13

[10] RLG-OCLC (2002) p.13

[11] UNESCO, *Guidelines for the Preservation of Digital Heritage* (March, 2003) p.43

sustainable business model to support its digital preservation/curation mandate.12 In the case of a state-funded TDR, the financing should be included in the formal planning documents. A private TDR should be able to guarantee its financial sustainability on the basis of the charged uses of its services and on a long-term business plan. DARIAH does recognise that there can be potential difficulties in securing long-term funding, especially for repositories that are funded by funding agencies.

Technological and procedural suitability

A TDR should consider and adopt the most appropriate preservation strategies and will communicate with its stakeholders about the suitability of various strategies. A TDR should ensure that it has the appropriate infrastructure for acquisition, storage and access. Repositories seeking DARIAH accreditation should refer to DARIAH policy Collection Ingest, Management and Preservation for guidance complying with DARIAH standards. In addition, the qualification and training of staff should be suitable for the goals, tasks and processes of the TDR. A TDR should also undergo regular external audits on systems components and performance.[13]

System Security

All the systems used in the operation of the TDR should be designed to assure security of systems for digital assets. A TDR should establish policies and procedures to meet requirements. These policies should meet community expectations, particularly those relating to copying processes, required redundancy of data, authentication systems, firewalls and backup systems.[14]

Procedural accountability

A TDR is responsible for a range of interrelated tasks and functions; therefore, it is accountable for all relevant policies and procedures. Documentation assures all stakeholders that a repository is meeting its requirements. These can include preservation policies; preservation strategic plans; preservation implementation plans; policies for collections development; policies for access control; policies that define the repository's designated community and policies for reviews, surveys and feedback. These policies and procedures should be documented and made available on request. A TDR must establish monitoring

[12] UNESCO, *Guidelines for the Preservation of Digital Heritage* (March, 2003) pp.43-44.

[13] RLG-OCLC (2002) p. 14

[14] RLG-OCLC (2002) p.14

mechanisms to ensure the continued operation of systems and procedures. It will also establish feedback mechanisms for problem resolution and to negotiate evolving requirements between providers and consumers.[15]

Roles and Responsibilities of a TDR

High-Level Organisational and Curatorial Responsibilities

A repository needs to fully understand what responsibilities they should assume for the preservation of digital materials. In this context, organisational responsibility should be understood at three levels: (i) their own local requirements; (ii)which other organisations might share some of the responsibilities through geography or arrangements such as consortia agreements or shared user communities, disciplines or format of materials; and (iii) which responsibilities can be shared and how.[16] High-level organisational for a TDR include the following:

Scope of collections

The scope of digital materials for repositories now ranges from simple e.g. text-based, digital files to complex multimedia and database resources. This diversity of digital materials and the role that they can play in the collection make development and application of collection policies extremely challenging.[17] For materials that have a physical counterpart, preservation decisions take into account considerations such as the condition of the original materials and the reason for digitising. Born-digital materials can be more challenging as their 'being digital' is the only method of access.

Preservation and lifecycle management

Preservation decisions cannot be postponed until it has been proven that the materials are worth keeping. Failure to agree on an active preservation and lifecycle management strategy will result in preservation actions that are more labour intensive, more complex and more expensive.

[15] RLG-OCLC (2002) p.15

[16] RLG-OCLC (2002) p.17

[17] RLG-OCLC (2002) p.17

The wide-range of stakeholders

Potential stakeholders can range from content creators, systems developers, custodians and future users; this can complicate the determinants of responsibilities, i.e. who does what, when and for how long. Decisions about how the materials are handled when created or maintained determine how or whether the repository can preserve them.

Ownership of material and other legal issues

The ownership of digital materials is not always straightforward, as digital materials are less tangible than a book, for example. Traditionally, responsibility for preservation was considered in conjunction with ownership of the materials, that is, the owner of the materials was responsible for determining the life span of the materials.18 Whereas with the preservation of digital material, it may be necessary to perform actions on the digital objects in order to keep the object accessible and usable. The lines of responsibility can be blurred in this context. Repositories should investigate and disseminate information about the complex relationship between digital preservation, licensing agreements and intellectual property rights.

Cost implications

It is generally accepted that digital preservation will require continuous resource commitments. Traditional and digital preservation should be compared with caution because the complex dependencies between long-term maintenance and continuing access make comparison problematic. Digital preservation is likely to draw on resources longer than traditional preservation does. It may be the case that different technical strategies such as different types of migration or emulation will require different costing timeframes and schedules.[19] It is important that repositories understand where the main expenses are likely to fall and how within existing practices in order for these to be incorporated to achieve economies of scale.

Operational Responsibilities

The OAIS Reference Model is a useful framework for identifying the responsibilities of a TDR for a repository that wishes to be OAIS-complaint. The following list of responsibilities defines the principle obligations of an OAIS-compliant repository. A reliable digital repository will:

[18] RLG-OCLC (2002) p.18

[19] RLG-OCLC (2002) p.19

- Negotiate for Appropriate Information from Content Providers;
- Obtain Sufficient Control of the Information;
- Determine the Repository's Designated Community;
- Ensure the Information to be preserved is Independently Understandable to the Designated Community;
- Follow Documented Policies and Procedures;
- Make the Preserved Information Available to the Designated Community; and
- Advocate Good Practice in the Creation of Digital Resources.

Certification

Certification is the process of assessing the degree to which a preservation programme complies with an agreed set of minimum standards or practices. A process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information. For creators, it provides a means of quality assurance when choosing potential service providers. For repositories, it supports repositories that need to provide objective evidence. A TDR can demonstrate its commitment to a regular schedule of self-assessment and certification in a number of ways including completed, dated checklists from self-assessments and/or third-party audits; certificates awarded for compliance with relevant ISO standards; timetables and evidence of adequate budget allocations for future certification. There are no widely-accepted standards for trust and there is no ISO standard for certification of trusted digital repositories. There are tools for TDR audit and certification such as checklist-based standards (nestor, TRAC, Data Seal of Approval) and flexible toolkits (DRAMBORA). Both TRAC and DRAMBORA are seeking to become recognised standards. A repository that wishes to comply with DARIAH requirements may use the tools and ISO standards outlined below as a tool for objective evaluation.

There are number of standards and documents that may be used as complementary audit tools such as:

- ISO 9000 family of standards addresses quality assurance components within an organisation and system management. These standards were not specifically developed to gauge the trustworthiness of a digital repository;

- ISO 17799:2005 was developed to address data security and information management. Like the previous standard it was not developed to address the trustworthiness of a digital repository;

- ISO 15489-1:2001 and ISO 15489-2:2001 both these standards define a systematic and process-driven approach, which governs the practice of records managers. ISO15489 requires an organisation to establish and document policies, practices and procedures for records management;

- ISO 14721:2003 is the OAIS Reference Model Standard and it provides a high-level reference model or framework identifying participants in digital preservation, their roles and responsibilities. Institutions can declare themselves 'OAIS-compliance' in order to underscore the trustworthiness of their repository. However, there is no established understanding of 'OAIS-compliance' beyond meeting the high-level responsibilities defined by the standard.[20] The OAIS approach provides a tool to evaluate whether preservation is successful.

An institution that has undertaken any kind of certification process, even if none of the components overlap with a digital repository audit, will be more prepared for a digital repository certification. Furthermore, organisations that have been certified in related standards will be able to use those certifications as evidence during a digital repository audit.21 As stated above there are tools for TDR audit and certification such as checklist-based standards - nestor, TRAC, Data Seal of Approval - and flexible toolkits – DRAMBORA that organisations can use in the absence of an internationally recognised standard.

RLG/OCLC – Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC)

In 2007 Research Libraries Group and Online Computer Library Centre (RLG/OCLC) published Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC) as a follow-up to the 2002 publication, Trusted Digital Repositories: Attributes and Responsibilities. TRAC is a set of criteria applicable to a range of digital repositories and archives. It provides tools for the audit, assessment and potential certification of digital repositories. It lists ninety organisational, technological and digital object management criteria for digital repositories. The criteria are not prescriptive; they are a recommendation

[20] Consultative Committee for Space Data Systems (CCSDS), *Audit and Certification of Trustworthy Digital Repositories: Draft Recommended Practice* (October 2009), p. 2-2

[21] CCSDS (2009) p.2-2

for the development of a certification programme for digital repositories. This checklist is one mechanism a repository can use to understand its capabilities, where it stands against potential threats and any other risks inherent in its system.[22] The aim of this initiative is turn these guidelines into a clear and recognised ISO standard which then can be used as a certification and auditing tool in digital preservation. For more information please see: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Network of Expertise in Long-Term Storage of Digital Resources (nestor)

In Germany nestor has designed a catalogue of fourteen criteria for TDR for long-term preservation. The overall aim is to introduce stable criteria for a wide variety of long-term digital repositories and to maintain the criteria over a long period. To this end, the fourteen catalogue criteria have been formulated at an abstract level. Each criterion is accompanied by detailed explanations and concrete examples. It is organised into the following sections: Organisation Framework, Object Management and Infrastructure and Security. The catalogue was also designed to conform to the OAIS Reference Model. For more information please see http://www.langzeitarchivierung.de/eng/index.htm

Data Seal of Approval

The Data Seal of Approval (DSA) was established by a number of institutions committed to the long-term archiving of research data. The DSA is granted to repositories that are committed to archiving and providing access to scholarly data in a sustainable way. In order to achieve the DSA and receive permission to use its logo, a repository must keep a file directory on the web that is accessible through the front-page of the repository. The completion of this self-assessment form is the starting point for the reviewing procedure. The assessment lists sixteen guidelines and the organisation describes how the guidelines relate to the repository and how they have been interpreted. The DSA Board reviews the form in order to decide if the repository will be granted the Seal of Approval. There is no audit, no certification, just a review on the basis of trust. Once approval has been granted, the DSA logo can be used by the repository. For more information please see www.datasealofapproval.org

[22] CCSDS (2009) p.2-2

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)

This tool presents a methodology of self-assessment, encouraging organisations to establish a comprehensive self-awareness of their objectives, activities and assets before identifying, assessing and managing the risks implicit within their organisation. Digital curation is characterised as a risk-management activity. DRAMBORA is also working towards becoming an ISO standard for risk assessment of record management. For more information please see http://www.repositoryaudit.eu/

DRAMBORA, TRAC and nestor are co-operating closely together and as a result of this co-operation, the ten principles of trust were formulated in 2007 as a leading principle for trustworthy repositories. 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:[23]

- The repository commits to continuing maintenance of digital objects for identified community/communities.
- Demonstrates organisational fitness (including financial, staffing structure, and processes) to fulfill its commitment.
- Acquires and maintains requisite contractual and legal rights and fulfils responsibilities.
- Has an effective and efficient policy framework.
- Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.
- Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
- Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.
- Fulfils requisite dissemination requirements.

[23] http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re accessed 4 June 2010

- Has a strategic programme for preservation planning and action.
- Has technical infrastructure adequate to continuing maintenance and security of its digital objects.