

# APPENDIX E: GUIDELINES FOR EVALUATING CLEARINGHOUSES

JUNE 2016



## APPENDIX E: GUIDELINES FOR EVALUATING DATA CLEARINGHOUSES

### Appendix E.1: Standard Guidelines “Data Seal of Approval” (1-16)

#### 0. Repository/Clearinghouse context

Guidance: To assess a repository/clearinghouse, reviewers need some information about the repository to set it in context.

- » Type
- » Designated Community
- » Level of Curation
- » Outsource Partners

*1. The clearinghouse has an explicit mission to provide access to and preserve data in its domain.*

Owner/Developer of the clearinghouse take responsibility for stewardship of digital objects, and to ensure that materials are held in the appropriate environment for appropriate periods of time.

*2. The clearinghouse maintains all applicable licenses covering data access and use and monitors compliance.*

Owner/Developer of the clearinghouse must maintain all applicable licenses covering data access and use, communicate about them with users, and monitor compliance. This requirement relates to the access regulations and applicable licenses set by the data repository itself, as well as any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information.

*3. The clearinghouse has a continuity plan to ensure ongoing access to and preservation of its holdings.*

The level of responsibility undertaken for data holdings, including any guaranteed preservation periods.

*4. The clearinghouse ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.*

Adherence to ethical norms is critical to responsible science. Disclosure risk—for example, the risk that an individual who participated in a survey can be identified or that the precise location of an endangered species can be pinpointed—is a concern that many repositories must address.

*5. The clearinghouse has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.*

Repositories need funding to carry out their responsibilities, along with a competent staff who have expertise in data archiving.

*6. The clearinghouse adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).*

An effective clearinghouse strives to accommodate evolutions in data types, data volumes, and data rates, as well as to adopt the most effective new technologies in order to remain valuable to its Designated Community. Given the rapid pace of change in the research data environment, it is therefore advisable for a repository to secure the advice and feedback of expert users on a regular basis to ensure its continued relevance and improvement. For this Requirement, responses should include evidence related to the following questions: Does the repository have in-house advisers, or an external advisory committee that might be populated with technical members, data science experts, and disciplinary experts?

*7. The clearinghouse guarantees the integrity and authenticity of the data.*

The clearinghouse should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access. Integrity ensures that changes to data and metadata are documented and can be traced to the rationale and originator of the change. Authenticity covers the degree of reliability of the original deposited data and its provenance, including the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

*8. The clearinghouse accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.*

The appraisal function is critical in determining whether data meet all criteria for inclusion in the collection and in establishing appropriate management for their preservation. Care must be taken to ensure that the data are relevant and understandable to the Designated Community served by the clearinghouse.

*9. The clearinghouse applies documented processes and procedures in managing archival storage of the data.*

Clearinghouses need to store data and metadata from the point of deposit, through the ingest process, to the point of access. Clearinghouses with a preservation remit must also offer 'archival storage' in OAIS terms.

*10. The clearinghouse assumes responsibility for long-term preservation and manages this function in a planned and documented way.*

The clearinghouse, data depositors, and Designated Community need to understand the level of responsibility undertaken for each deposited item in the repository. The repository must have the legal rights to undertake these responsibilities.

*11. The clearinghouse has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.*

Clearinghouses must work in concert with depositors to ensure that there is enough available information about the data such that the Designated Community can assess the substantive quality of the data. Clearinghouses must also be able to evaluate the technical quality of data deposits in terms of the completeness and quality of the materials provided, and the quality of the metadata.

*12. Archiving takes place according to defined workflows from ingest to dissemination.*

To ensure the consistency of practices across datasets and services and to avoid ad hoc and reactive activities, archival workflows should be documented, and provisions for managed change should be in place. The procedure should be adapted to the clearinghouse mission and activities, and procedural documentation for archiving data should be clear.

*13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.*

Effective data discovery is key to data sharing, and most repositories provide searchable catalogues describing their holdings such that potential users can evaluate data to see if they meet their needs. Once discovered, datasets should be reference able through full citations to the data, including persistent identifiers to ensure that data can be accessed into the future. Citations also provide credit and attribution to individuals who contributed to the creation of the dataset.

14. *The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.*

Clearinghouses must ensure that data can be understood and used effectively into the future despite changes in technology. This requirement evaluates the measures taken to ensure that data are reusable.

15. *The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.*

Clearinghouses need to operate on reliable and stable core infrastructures that maximizes service availability. Furthermore, hardware and software used must be relevant and appropriate to the Designated Community and to the functions that a clearinghouse fulfils. Standards such as the OAIS reference model specify the functions of a clearinghouse in meeting user needs.

16. *The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.*

## **Appendix E.2: Guidelines for Other Evaluation Criteria**

### NESTOR Seal for Trustworthy Digital Archives

- » C1: Selection of information objects and their representations
  - » Criteria have been defined for the selection of information objects and their representations in the digital archive. The framework is provided by legal obligations, the institution's or companies basic function, and its own targets.
- » C2: Responsibility for preservation
  - » The digital archive assumes responsibility for the long-term preservation of the information objects on the basis of legal requirements or its own objectives. Long-term preservation means ensuring the long term usability of the information contained in the representations.
- » C3: Designated communities
  - » The digital archive has defined its designated community/communities. This includes knowledge of the specific requirements of the designated communities which influence the selection of the services to be provided. If the designated communities or their requirements change over time, the digital archive should adapt accordingly.
- » C4: Access
  - » The digital archive ensures that authorized users in the designated communities can access the representations. This includes appropriate search possibilities. The digital archive openly declares its conditions of use and any costs which may arise, listing these in a transparent manner.
- » C5: Interpretability
  - » The digital archive has defined measures to ensure the long-term interpretability of at least one of the representations, thereby meeting a basic precondition for appropriate use now and in the future. This includes the interpretability of both content data and metadata. In ensuring this, the digital archive should take the needs of its designated community/communities into account. Changes to the technical environment or the designated community or communities can influence the interpretability of the objects. Using appropriate procedures, the digital archive should therefore check at regular

intervals whether the objects can still be interpreted by the designated community or communities.

- » C6: Legal and contractual basis
  - » The digital archives ingest, archiving and access procedures are based on legal or contractual regulations concluded with the producers. The nature and scope of the delivery is regulated, as are the digital archive's archival obligations, the conditions of use and, where applicable, the costs.
- » C7: Legal conformity
  - » The digital archive monitors and documents conformity with relevant regulations concerning the ingest, archiving and use of digital objects. These include: data protection, protection of the rights of affected parties, confidentiality regulations, copyright and usage rights, internal and external compliance.
- » C8: Funding
  - » Valid budget planning exists, as does a long-term funding plan for the digital archive.
- » C9: Personnel
  - » Sufficient numbers of appropriately qualified staff are available. Updated job descriptions exist which set out the required qualifications of the digital archive personnel and contain an organizational chart and/or a staff development plan based on the tasks and objectives of the digital archive.
- » C10: Organization and processes
  - » The organizational structure should be appropriate for the objectives, tasks and processes of the digital archive. The structural and procedural organization should be defined. The responsibilities should be established. The digital archive is incorporated at the appropriate point in the schedule of responsibilities.
- » C11: Preservation measures
  - » The digital archive should conduct strategic planning as a means of preserving the digital objects entrusted to it. This should include imminent or expected tasks and specify the deadlines by which they are to be completed. Long-term planning should be based on the monitoring of legal and social changes, the demands and expectations of the designated communities and all technical changes relevant for the sustained preservation and appropriate use of the information objects in the form of their representations. Possible effects on task fulfilment are evaluated. Suitable structures and procedures exist for this.
- » C12: Crisis / successorship management
  - » The digital archive is in possession of a plan which ensures continuation of the preservation tasks even beyond the archive's own existence. The digital archive should have made contingency plans. In such a case the preservation work must be continued in a different organizational framework, thereby ensuring that the set tasks can be carried out in full. Where this is not possible, any deficiencies should be documented. The digital archive should take precautions to ensure that the transition process can be defined, planned and implemented in good time.
- » C13: Significant properties
  - » The digital archive identifies and documents which of the transferred representations' properties are significant for preservation of the information objects. In determining the scope of the properties to be preserved, a balance should be struck, bearing the archive's own targets in mind, between the technical possibilities and the costs of long-term

preservation on the one hand and the needs of the designated community/communities on the other hand.

- » C14: Integrity: Ingest interface
  - » The digital archive has its own interface for ingesting the representations in a way which retains their integrity. The interface contains all of the functions and processes aimed at transferring the submission information packages from the producers, transforming them into archival information packages and incorporating them into the digital archive. The interface allows the producer and the digital archive administration to check and maintain the integrity of the representations.
- » C15: Integrity: Functions of the archival storage
  - » The archival storage provides functions necessary for checking and maintaining the integrity of the representations by the administration of the digital archive. The functions include recording of the archival information packages onto storage media, long-term storage, and restoration of the archival information packages and all changes to the packages.
- » C16: Integrity: user interface
  - » The digital archive has an interface which allows users and the digital archive administration to check and maintain the integrity of the representations. This includes the transformation from archival information packages into dissemination information packages.
- » C17: Authenticity: Ingest
  - » The digital archive has procedures which permit the authenticity of the representations to be assessed upon being ingested and the authenticity of the submission information packages to be assessed and protected.
- » C18: Authenticity: Preservation measures
  - » The digital archive deploys methods which ensure the authenticity of the objects during implementation of the long-term preservation measures and document the degree of authenticity.
- » C19: Authenticity: Use
  - » The digital archive allows the users and the administration of the digital archive to check and maintain the authenticity of the representations. This includes the transformation from archival information packages into dissemination information packages.
- » C20: Technical authority
  - » The digital archive obtains technical authority over the representations being ingested, allowing it to transform them into archival information packages and, if necessary, to carry out long-term preservation measures. Following the transfer, all necessary measures can be carried out without any technical restrictions.
- » C21: Submission information packages
  - » The digital archive has issued specifications regarding its submission information packages. The digital archive agrees with the producers which submission information packages will be ingested (content data and metadata). The submission information packages are checked on the basis of the specifications.
- » C22: Transformation of the submission information packages into archival information packages
  - » The digital archive converts submission information packages into archival information packages.

- » C23: Archival information packages
  - » The digital archive has issued specifications for its archival information packages. The digital archive defines which archival information packages (content data and metadata) are to be stored and in which form. The archival information packages are checked on the basis of the specifications.
- » C24: Interpretability of the archival information packages
  - » Technical preservation measures are undertaken to ensure the interpretability of the archival information packages.
- » C25: Transformation of archival information packages into dissemination information packages
  - » The digital archive transforms archival information packages into dissemination information packages.
- » C26: Dissemination information packages
  - » The digital archive specifies the dissemination information packages on the basis of the requirements of the designated communities.
- » C27: Identification
  - » A digital archive should use internal identifiers to manage the information objects and their representations and, where applicable, their parts and relationships (part/totality, different variants, versions etc.), especially to ensure unique assignment of the content data to the metadata. The use of externally visible, standardized persistent identifiers ensures reliable tracing of the information objects and their representations, and consequently also access.
- » C28: Descriptive metadata
  - » The scope, structure and content of the descriptive metadata are defined. They depend on the goals of the digital archive, its designated communities and the object types.
- » C29: Structural metadata
  - » The structure of the representations must be adequately described so that the information objects can be reconstructed and used.
- » C30: Technical metadata
  - » The technical metadata are defined to ensure interpretability, integrity and authenticity and to manage the preservation measures.
- » C31: Logging the preservation measures
  - » The digital archive logs the preservation measures and any changes to the representations.
- » C32: Administrative metadata
  - » The digital archive has defined its administrative metadata in order to render the administration and use of the information objects and their representations transparent. Use of the representations may be restricted for legal or contractual reasons.
- » C33: IT infrastructure
  - » The IT infrastructure should realize the specifications for handling the information objects and representations at the technology and security levels.
- » C34: Security
  - » The organization and the infrastructure protect the digital archive and its archived information objects and representations.

## ISO16363: Audit and Certification of Trustworthy Digital Repositories

| 3. ORGANIZATIONAL INFRASTRUCTURE                              |       |   |
|---|-------|---|
| 3.1 GOVERNANCE & ORGANIZATIONAL VIABILITY                     |       |   |
|   |       | Metric  |
|   | 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 & STAFFING                       |       |   |
|   |       | Metric  |
|   | 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 FULFIL THESE DUTIES.   |
| 3.3 PROCEDURAL ACCOUNTABILITY & PRESERVATION POLICY FRAMEWORK |       |   |
|   |       | Metric  |
|   | 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                                  |       |   |
|   |       | Metric  |
|   | 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. |