CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2020–2022

This Glossary of Terms supports the CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022: https://doi.org/10.5281/zenodo.3638211.

(Taken from: * OAIS[1]; † the Society of American Archivists[2]; ‡ the CASRAI Dictionary[3])

Access Rights Information*: The information that identifies the access restrictions pertaining to the content information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the submission agreement, related to both preservation (by the repository) and final usage (by the consumer). It also includes the specifications for the application of rights enforcement measures.

Appraisal†: The process of determining whether records and other materials have permanent (archival) value. Appraisal may be done at the collection, creator, series, file, or item level. Appraisal can take place prior to donation and prior to physical transfer, at or after accessioning.

Archive: An organization, place, or collection that stores static digital and analogue records, documents, and other materials for long-term preservation such that it can be accessed and reused by a Designated Community.

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.

Consumer*: The role played by those persons, or client systems, who interact with [repository] services to find preserved information of interest and to access that information in detail. This can include other repositories, as well as internal repository persons or systems.

Curation‡: The activity of managing and promoting the use of data from their point of creation to ensure that they are fit for contemporary purpose and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep them fit for purpose. Higher levels of curation will also involve links with annotation and with other published materials.

Data*: A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.

Database‡: A collection of data that is organized according to a conceptual structure/model describing the characteristics of these data and the relationships among their corresponding entities, supporting one or more application areas. A database allows its contents to be easily accessed, managed and updated...A database may house one or many datasets.

Dataset‡: Any organized collection of data in a computational format, defined by a theme or category that reflects what is being measured/observed/monitored. The presentation of the data in the application is enabled through metadata.

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 [repository] and this definition may change over time.

Digital Object*: An object composed of a set of bit sequences.

Digital Preservation‡: The series of managed activities necessary to ensure continued access to digital materials for as long as necessary. Digital preservation is defined very broadly and refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change. Those materials may be records created during the day-to-day business of an organization; 'born-digital'

---

[1] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Recommended Practice -- CCSDS 650.0-M-2. Magenta Book, June 2012.http://public.ccsds.org/publications/archive/650x0m2.pdf

[2] https://www2.archivists.org/glossary/terms

[3] https://dictionary.casrai.org

materials created for a specific purpose (e.g., teaching resources); or the products of digitization projects. This definition specifically excludes the potential use of digital technology to preserve the original artefacts through digitization.

Domain or Subject-based Repository: Specializes in a specific research field or data type.

Ingest: The process of entering data and associated metadata into a data repository.

Integrity: Internal consistency or lack of corruption of digital objects. Integrity can be compromised by hardware errors even when digital objects are not touched, or by software or human errors when they are transferred or processed.

Institutional Repository: Repository serving a research performing institution. Often are generic, multi-subject repositories; for example, within a university.

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

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 a repository. This period extends into the indefinite future.

Long-term Preservation‡: Continued access to digital materials, or at least to the information contained in them, indefinitely.

Migration‡: A means of overcoming technological obsolescence by transferring digital resources from one hardware/software generation to the next. The purpose of migration is to preserve the intellectual content of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration differs from the refreshing of storage media in that it is not always possible to make an exact digital copy or replicate original features and appearance and still maintain the compatibility of the resource with the new generation of technology.

National Repository System, Including Governmental: Multidisciplinary, national infrastructure. Has a legal mandate for certain (public or governmental) data types.

Open Archival Information System (OAIS)*: [A]n organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS Archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this recommendation [i.e., the Reference Model] and future related recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

Preferred Formats: Formats that a repository can reasonably assure will remain readable and usable. Typically, these are the de facto standards employed by a particular discipline.

Producer*: The role played by those persons or client systems that provide the information to be preserved. This can include other repositories or internal repository persons or systems.

Provenance Information*: The information that documents the history of the content information. This information tells the origin or source of the content information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. The [repository] is responsible for creating and preserving provenance information from the point of ingest; however, earlier provenance information should be provided by the producer. Provenance information adds to the evidence to support authenticity.

Publication Repository: A repository that collects scholarly or other textual data linked to publications, including pre- and postprints.

Reference Model*: A framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist.

Repository‡: [Organizations that] preserve, manage, and provide access to many types of digital materials in a variety of formats. Materials in online repositories are curated to enable search, discovery, and reuse. There must be sufficient control for the digital material to be authentic, reliable, accessible and usable on a continuing basis.

Research Project Repository: Capture research results that require a deposit mandate by a funder or organization as a 'record of science'. Often tied to a specific (multi)disciplinary project.

Reuse: The use of data collected for one purpose to study a new problem or to verify the conclusions of the data producer.

Stewardship‡: A shared responsibility between Principal Investigators and data stewards. Principal Investigators are responsible for, and data stewards provide support for: (a) Data collection, data integration, or reuse of existing data; (b) Review of data quality; (c) Description of scientific workflow/process; (d) Provision of standards-compliant metadata; and, (e) Submission of data and data productions. Data stewards are responsible for, and Principle Investigators are consulted and informed on: (a) Preservation of data and data products; and, (b) Provision of formats (e.g., web services, NetCDF, etc.) for data discovery and integration. In addition, Principal Investigators are also responsible for data citation, as appropriate, when preparing documentation, reports, or references.

Succession Plan*: The plan of how and when the management, ownership and/or control of the repository holdings will be transferred to a subsequent repository in order to ensure the continued effective preservation of those holdings.