CoreTrustSeal Trustworthy Data Repositories Requirements: Glossary 2023-2025

v01.00

This Glossary of Terms from other sources is provided to support the CoreTrustSeal Requirements and Extended Guidance.

The CoreTrustSeal texts use 'Digital Objects' to refer to 'data', 'metadata' and other materials managed together by repositories. These encompass the definitions provided for metadata, datasets and digital objects provided below.

OAIS uses the term Producer (included below for context alongside other definitions). The CoreTrustSeal texts use 'depositor' to refer to the actor that offers a digital object to a repository to be curated and preserved because that actor may not be the literal 'producer' of the data or metadata.

Definitions for Archive and Repository are provided below. The CoreTrustSeal texts use the term repository, but also follow OAIS (Open Archival Information System) assumptions about mandatory responsibilities. Some organisations that self-identify as archives or repositories may not meet those mandatory responsibilities, including active preservation of digital objects, and so may not be in scope to become CoreTrustSeal- certified.

(Taken from: * OAIS[1]; † the Society of American Archivists[2]; ‡ the CASRAI Dictionary[3,][**] the DPC Handbook[4])

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.

Appraisal[]: Appraisal of born digital objects should include a measured assessment of their value to the parent organisation set against the challenges of long-term preservation and providing access. These challenges may include an organisation's ability to read or open a version of the master file, the ability to secure sufficient rights to manage and provide access to current and future versions of the file, or simply staffing and funding resources [...]. It should be remembered that organisations can provide access to resources that they have accessioned without placing them in specific preservation or retention workflows. A detailed policy document which clearly identifies the most important digital resources (from either a format or content perspective) can give guidance on appraisal of born digital objects destined

---

[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://codata.org/initiatives/data-science-and-stewardship/rdm-terminology-wg/rdm-terminology/ (currently under review)
[4] https://www.dpconline.org/handbook/glossary

for such pathways). [5]

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 digital material is what it purports to be. In the case of electronic records, it refers to the trustworthiness of the electronic record as a record. In the case of "born digital" and digitised materials, it refers to the fact that whatever is being cited is the same as it was when it was first created unless the accompanying metadata indicates any changes. Confidence in the authenticity of digital materials over time is particularly crucial owing to the ease with which alterations can be made.

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.

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.

Depositor: See Producer

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' 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.

Specialist Repository: Domain or Subject-based Repository which specializes in a specific (research) field or data type, and supports that defined designated community.

Generalist repository: a repository that does not specialise in a domain, discipline, specific

---

[5] https://www.dpconline.org/handbook/organisational-activities/acquisition-and-appraisal

(research) field or data type and supports a defined designated community.

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.

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.

Metadata**: Information which describes significant aspects of a resource.

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.

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 recommends or requires to be provided at deposit, or migrated to for preservation or access, to reduce the risk of them becoming less readable or usable over time. These may include the de facto standards employed by a particular discipline.

Preservation: See Digital Preservation

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.

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.

Reuse: The use of data or metadata collected for one purpose to study a new problem or to verify the conclusions of prior work

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.