

Curation & Preservation Levels¹

CoreTrustSeal Board Discussion Paper

Context	1
Introduction	2
Digital objects as the focus of active preservation	3
Draft Example of Tiered Curation and Preservation	4
Concluding Thoughts and Next Steps	4
Appendix: CoreTrustSeal v2.0 Levels of Curation	5

Context

Applying appropriate levels of curation and preservation to digital objects maximises the long term return on investment in data assets.

Curation and Long-term preservation depend on a repository having the rights, and taking the responsibility to provide an organisational infrastructure, digital object management and technical/security setting that is capable of supporting this service. The CoreTrustSeal Requirements (v2.0² and proposed v3.0³) request information about the levels of *curation* an applicant offers⁴. These reflect an assumption that applicants must take responsibility for active long term digital *preservation* for a designated community. As the issues of curation, preservation⁵ and certification are receiving more attention from a wider range of actors, the need for clearer specification of preservation levels has become clear. The CoreTrustSeal Board sees this as an important issue for the data management community, for defining which applicants are in-scope for certification, and for better defining 'non-preservation' data and metadata services⁶. In addition to revision of the CoreTrustSeal requirements⁷ there are a number of other open consultations such as the COAR Community Framework for Good

¹ Cite as: CoreTrustSeal Standards and Certification Board. (2022). Curation & Preservation Levels (v01.00). Zenodo. <https://doi.org/10.5281/zenodo.6908019>.

² CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022 <https://doi.org/10.5281/zenodo.3632533>

³ DRAFT Change Log and CoreTrustSeal Requirements 2023-2025 <https://doi.org/10.5281/zenodo.6669515>

⁴ See Appendix: CoreTrustSeal v2.0 Levels of Curation

⁵ <https://www.eosc.eu/advisory-groups/long-term-data-preservation>

⁶ CoreTrustSeal: Specialists, Generalists, and Technical Repository Service Providers <https://doi.org/10.5281/zenodo.3964071>

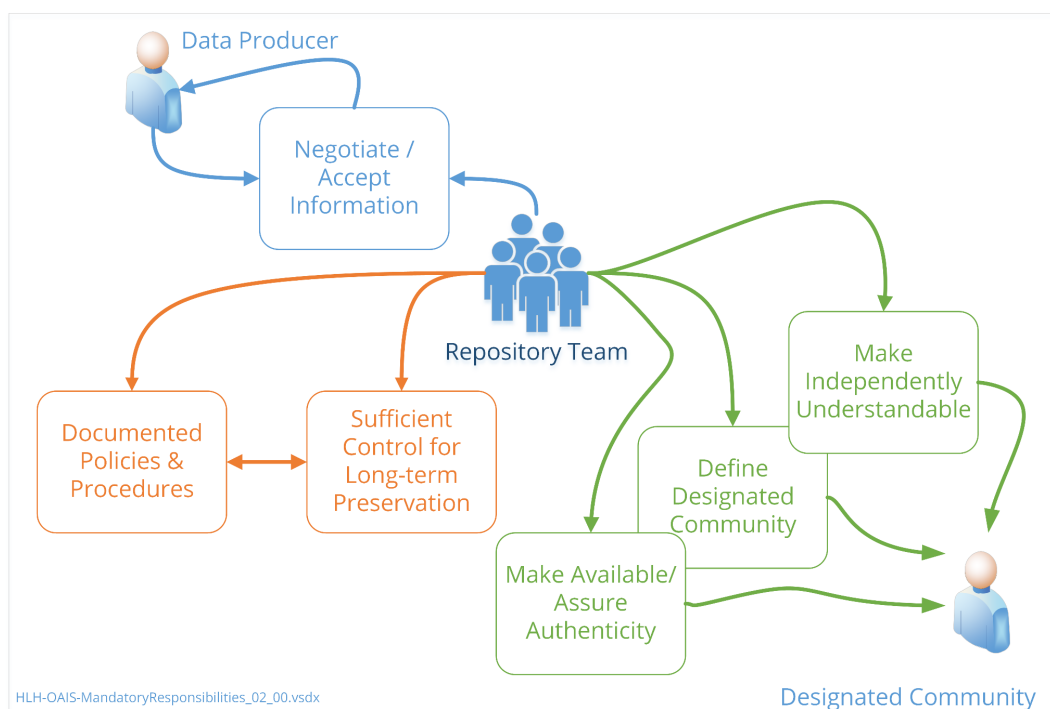
⁷ [Trustworthy Data Repository Requirements Review 2023-2025](#)

Practices in Repositories⁸ and the NDSA Levels of Preservation⁹. The CoreTrustSeal Board is sharing this draft discussion document for community comment as a first step towards reaching consensus.

Introduction

In addition to delivering 'core' level Trustworthy Digital Repository (TDR) certification the CoreTrustSeal seeks to align with and contribute to the wider data lifecycle and landscape. For disciplinary and generalist data repositories, and across the (meta)data product and service provider ecosystem, the level of curation and preservation delivered for each object must clearly be communicated to data users, and to other stakeholders, including policy makers and funders.

If curation involves actions that deliver an immediate benefit to digital objects, preservation can be seen as the measures that ensure data and metadata will remain accessible, usable and understandable by a designated community into the future. Preservation takes account of ongoing changes to the knowledge base of that community and the surrounding technical context. Long-term doesn't have to mean 'forever'. Objects may be reappraised over time to change their level of preservation. Long-term preservation means that organizational measures, infrastructure, and policies are in place to actively preserve for *as long as necessary*. Minimum periods of retention are important and should be clear, but these do not equate to active preservation. To qualify for the CoreTrustSeal a repository must deliver active preservation while meeting a number of mandatory responsibilities.



Mandatory Responsibilities¹⁰

⁸ <https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories/>

⁹ <https://ndsa.org/publications/levels-of-digital-preservation/>

¹⁰ See section 3.1 of Reference Model for an Open Archival Information System (OAIS), <https://public.ccsds.org/pubs/650x0m2.pdf>

To remain ‘understandable’ and in line with principles such as FAIR¹¹ (Findable, Accessible, Interoperable and Re-Usable) it is necessary to ensure that supporting metadata also remains fit for the purposes of designated community use and ongoing preservation. Both data and metadata need to be preserved through managed changes that address, for example, the knowledge or skill set of the designated community. Common approaches to active preservation thus include transformations to new data formats and metadata schemas, and updates to (meta)data content so that digital objects remain understandable and technically usable by the community.

The designated community’s needs and preferences must be taken into account when determining the preservation actions to be applied. This depends on monitoring the knowledge base and technology needs of the community, and an understanding of wider technical risks¹². For digital objects with specialist characteristics and users (e.g. disciplinary) the active preservation of (meta)data can be more challenging and require additional expertise. A more generalist approach may not preserve those characteristics or meet those specialist needs.

Digital objects as the focus of active preservation

Efforts to unite different perspectives on information management are not new¹³. Active digital preservation ensures the continued use and understanding of digital objects for a defined designated community.

“Every digital object is a physical object, a logical object, and a conceptual object, and its properties at each of those levels can be significantly different. A **physical** object is simply an inscription of signs on some physical medium. A **logical** object is an object that is recognized and processed by software. The **conceptual** object is the object as it is recognized and understood by a person, or in some cases recognized and processed by a computer application capable of executing business transactions”.¹⁴

The intellectual **conceptual** entity depends on a **logical** entity rendered through a given hard- and software environment, based on bits and bytes stored on a **physical** medium. To ensure the continued use and understanding of digital objects by a defined designated community a Trustworthy Digital Repository must provide active preservation at the physical, logical and conceptual level.

The draft levels presented below address conceptual and logical curation and preservation¹⁵. A single organisation may hold digital objects that are curated and preserved at different levels. These levels must be sufficiently specific and granular to communicate the care provided and the degree of responsibility taken by a repository or other data service at the object level.

¹¹ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

¹² Covered in detail in FAIR + Time: Preservation for a Designated Community <https://doi.org/10.5281/zenodo.4783116>
¹³ <https://www.zachman.com/ea-articles-reference/58-conceptual-logical-physical-it-is-simple-by-john-a-zachman>

¹⁴ https://chnm.gmu.edu/digitalhistory/links/pdf/preserving/8_37e.pdf ; see also <https://www.naa.gov.au/sites/default/files/2020-01/An-Approach-to-the-Preservation-of-Digital-Records.pdf> and OAIS for similar typologies.

¹⁵ All assume that effective physical measures are in place, including back-ups and multi-format, multi-location, multi-copy redundancy and integrity.

Draft Example of Tiered Curation and Preservation

All of the levels below are options in real-world appraisal decision-making. Levels Z and C are not in scope for CoreTrustSeal certification as they do not provide a long-term perspective. Agreement on these levels will support further discussion on how they should be applied and how supporting evidence should be provided.

Z. Level Zero. Content distributed as deposited. Unattended deposit-storage-access.

Data content and supporting metadata are distributed to users exactly as they are provided by depositors. No curation or long term preservation.

C. Basic Compliance and/or curation

Data content and supporting metadata deposited are checked at the point of deposit for compliance with defined criteria for data formats and metadata elements. If these criteria are not met the digital objects are returned to the depositor for change, or the repository undertakes the necessary curation steps to ensure they comply. Minimal curation for initial access and use, but no long term preservation.

B. Logical-Technical Curation

In addition to C above the repository takes long-term responsibility for ensuring that the data and metadata are updated over time to newer standards and formats in response to:

- i. technical risks (e.g. file format obsolescence) and/or
- li. the changing needs of the designated community (e.g. newer alternate formats become necessary for reuse).

A. Conceptual preservation for understanding and reuse

In addition to B and C above the repository monitors changes to the definition and demands of their designated community, including their knowledge base, and takes responsibility for the preservation actions that ensure digital objects can be understood and re-used.. Usually this will involve updates to the content of metadata elements and other semantic artifacts such as controlled vocabularies and ontologies. For some repositories it may include responsibility for editing the structure and content of deposited data.

Concluding Thoughts and Next Steps

This initial discussion paper is being shared with the community for comment with a view to future iterations and the development of a CoreTrustSeal position paper that may influence future versions of the Requirements. For version 3.0 of the Requirements for 2022-2024 the current levels of curation (see

Appendix) will be retained. In the future integrated curation and preservation levels that have been agreed by the community would provide a valuable reference point for communicating the degree of care a digital object receives and which actors take responsibility for that care. This would then provide insights into how those offering different levels of curation could be assessed and evaluated.

Appendix: CoreTrustSeal v2.0 Levels of Curation

"Level of Curation Performed. Select all relevant types from:

- A. Content distributed as deposited
- B. Basic curation – e.g., brief checking, addition of basic metadata or documentation
- C. Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- D. Data-level curation – as in C above, but with additional editing of deposited data for accuracy"