



Curation & Preservation Levels¹

CoreTrustSeal Board Position Paper v3.0

Context	1
Introduction	2
Digital Objects as the Focus of Active Preservation	3
Curation and Preservation Levels (v3.0)	4
Z. Level Zero. Content distributed as deposited. Unattended deposit-storage-access.	4
D. Deposit Compliance	4
C. Initial Curation	4
A. Active preservation	4
Concluding Thoughts and Next Steps	5
Appendix: CoreTrustSeal 2023-2025 Levels of Curation	5

Context

Applying the appropriate levels of care to digital objects maximises the return on investment in data assets over time. Successful curation and long-term preservation activities depend on a repository having the rights and taking the responsibility to provide an effective organisational infrastructure, digital object management and technical/security environment. Clearly communicating the levels of *curation* a repository offers² is an essential part of seeking certification against the CoreTrustSeal Requirements 2023-2025³. To be in scope for CoreTrustSeal applicants must also take responsibility for *active long-term digital preservation* for a defined ('designated') community of users. As the issues of curation, preservation⁴ and certification are receiving more attention from a wider range of actors, the need for clearer specification of curation and preservation levels has become obvious. The CoreTrustSeal Board sees this as an important issue for the data management community, for defining which applicants are in-scope for certification, and as a step toward the improved definition of all data and metadata services, including those that do not offer active preservation.⁵

Version 1.0 of this paper was published during the 2022 revision of the CoreTrustSeal Requirements. Version 2.0 addressed the feedback received from the community and was made available for further discussion. This version 3.0 has been approved by the Board as a position paper. It joins the previous

¹ Cite as: CoreTrustSeal Standards and Certification Board. (2024). Curation & Preservation Levels (v03.00). Zenodo. <https://doi.org/10.5281/zenodo.11476980>.

² See Appendix: CoreTrustSeal 2023-2025 Levels of Curation

³ CoreTrustSeal Standards and Certification Board. (2022). CoreTrustSeal Requirements 2023-2025 (V01.00). Zenodo. <https://doi.org/10.5281/zenodo.7051012>

⁴ EOSC Association. Long Term Data Preservation Task Force (LTDP-TF) Final Report & Recommendations. Zenodo. <https://doi.org/10.5281/zenodo.10820893>

⁵ CoreTrustSeal: Specialists, Generalists, and Technical Repository Service Providers <https://doi.org/10.5281/zenodo.3964071>

Logical-Technical and Conceptual (Semantic) levels into a single level A: Active preservation. In addition this version acknowledges that wider quality measures are also a consideration for preservation, and that logical, semantic and quality aspects are relevant to deposit compliance, initial curation, and active preservation.

Introduction

In addition to delivering Trustworthy Digital Repository (TDR) certification at a 'core' level, 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 can be understood as the actions that deliver an *immediate* benefit to digital objects, then preservation includes these, and other steps to ensure data and metadata remain accessible, usable and understandable *into the future*. Preservation takes account of ongoing changes to the landscape, e.g. the knowledge base of the user community, the surrounding technical context, and the ethical and legal environment. Long-term does not have to mean 'forever'. Objects may be reappraised over time and their level of curation or preservation may change. Long-term preservation means that organizational measures, infrastructure, and policies are in place to actively preserve digital assets 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. To remain 'understandable' and in line with principles such as FAIR⁶ (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 evolutions in the landscape. Common approaches to active preservation can 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 considered when determining the curation and preservation actions to be applied. This depends on monitoring the knowledge base and technology needs of the community, and an understanding of wider technical risks⁷ as well as technical, ethical, legal and other developments that may impact how digital objects can be used. 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

Active digital preservation ensures the continued use and understanding of digital objects for a defined designated community.

⁶ 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>

⁷ Covered in detail in FAIR + Time: Preservation for a Designated Community <https://doi.org/10.5281/zenodo.4783116>

“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”.⁸

The **semantic (conceptual)** entity depends on a **logical** entity rendered through a given hard- and software technical 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 by addressing the logical and/or semantic entity alongside the physical.

The proposed level A: Active preservation presented below comprises semantic and logical preservation. Separating semantic and logical preservation into distinct levels may be helpful in contexts where it is necessary to specify and communicate the care provided and the degree of responsibility taken at the object level in a more granular fashion.⁹ However, for the current purpose - facilitating the decision of a given applicant is in scope for CoreTrustSeal certification, it is not necessary.

Curation and Preservation Levels (v3.0)

The levels may be cumulative as they progress from Deposit Compliance (D), through Initial Curation (C), to Active Preservation (A). A repository may offer different levels of curation and preservation for different digital objects.

In addition to describing the types of care offered by an organisation, the levels enable repositories to describe and document the level of care at individual object level, thereby contributing to documented audit and provenance trails.

From the perspective of the CoreTrustSeal, Levels D and C alone are not in scope for CoreTrustSeal certification as they do not entail active long-term preservation and hence do not provide a long-term perspective beyond bit storage.

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

Data content and supporting metadata are stored for a given time period, or indefinitely. This may include multiple copies and monitoring of bitstreams for integrity. Data content and supporting metadata are distributed to users exactly as they are provided by depositors. Beyond these measures, there are no checks of deposit compliance, no initial curation or active long-term preservation.

⁸ 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.

⁹ All assume that effective physical measures are in place, including back-ups and multi-format, multi-location, multi-copy redundancy and integrity.

D. Deposit Compliance

Data content and supporting metadata deposited are checked for compliance with defined criteria, e.g. data formats, metadata elements, and compliance with legal and ethical norms.¹⁰ Digital objects that do not meet these criteria may be rejected, or moved forwards to initial curation if provided by the repository.

C. Initial Curation

The digital objects are curated by the repository to meet defined criteria, which may exceed those defined for Deposit Compliance. This initial curation for access and use may include, e.g., the correction or enhancement of metadata and/or data content, or the creation of dissemination formats.

A. Active preservation

In addition to D and/or C above the repository takes long-term responsibility for ensuring that the data and metadata can be understood and rendered as required by the designated community for reuse. The preservation actions can be aimed at logical-technical, semantic, or quality aspects of the (meta)data, for example, in response to the threat of technological obsolescence, to accommodate changing needs of the Designated Community, or in response to other considerations such as security or legal concerns.

Logical-technical measures include updating hard- and software environments, archival and dissemination formats of digital objects, and metadata.

Semantic measures include updating the content of metadata elements and other semantic artefacts such as controlled vocabularies and ontologies if necessary. It may include responsibility for editing the structure and content of deposited data.

Concluding Thoughts and Next Steps

This Board-approved Position Paper is a revised version of a discussion paper shared with the community for comment in 2023¹¹. It is published with a view to providing input into future versions of the Requirements. For the 2023-2025 version of the Requirements the current levels of curation (see Appendix) have been 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, and what supporting evidence should be provided at each level.

¹⁰ The actions that follow these checks are determined by the repository. For example, a repository may choose to return (meta)data that does not meet the deposit criteria to the depositor, or to ingest the (meta)data and document non-compliance, or to undertake initial curation to ensure compliance.

¹¹ CoreTrustSeal Standards and Certification Board. (2023). Curation & Preservation Levels: CoreTrustSeal Discussion Paper (v02.00). Zenodo. <https://doi.org/10.5281/zenodo.8083359>.

Appendix: CoreTrustSeal 2023-2025 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”