CoreTrustSeal Levels of Curation and Preservation: Implied Repository and Object Metadata Characteristics

This initial draft working paper is shared by the authors to seek feedback from across the research data management community.

The CoreTrustSeal Board proposed, and developed through open public feedback, the Curation and Preservation Levels[1]. The discussion paper was prompted by applications for certification from an increasingly broad range of organisations that hold digital objects. The resultant position paper defined the following levels[2] of care:

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

D. Deposit Compliance. Non-compliance triggers rejection, or requires initial curation.

C. Initial Curation. Repository takes action as required to meet defined criteria.

A. Active preservation. Long-term responsibility to take action as required to ensure reuse.

To be in scope for CoreTrustSeal certification a repository *must* offer active preservation. It *may* also offer deposit compliance and initial curation.

Criteria set and actions taken by the repository may be intended to address the logical (e.g. technical formats), semantic (e.g. metadata, ontologies) or quality aspects of digital objects.

---

[1] CoreTrustSeal Standards and Certification Board. (2024). Curation & Preservation Levels: CoreTrustSeal Position Paper. Zenodo. https://doi.org/10.5281/zenodo.11476980

[2] See Appendix B for the full descriptions.

Logical, semantic and quality issues inform the key intervention points where the repository takes action on digital objects: deposit compliance, initial curation, active preservation and reappraisal.

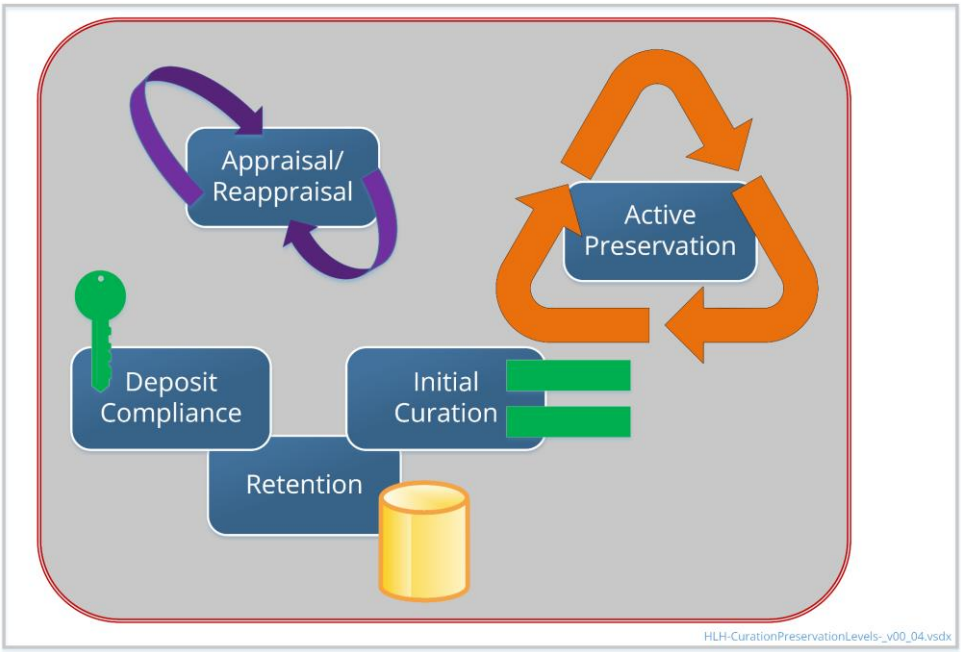

Diagram: Retention, Deposit, Curation, Preservation, ReAppraisal

Feedback to the discussion paper noted that the levels "could be applied across repositories and data services that offer everything from basic storage to full, disciplinary, active long term preservation. All of these types of service are necessary and form part of current research infrastructure, even if the CoreTrustSeal does not currently offer a certification solution for all of these levels."[3] However, the UK Data Service also noted that the levels themselves do not specify how a specific object is being cared for, or what information and artefacts should be shared to clarify the approach to deposit compliance, initial curation and preservation. In addition to specifying any guaranteed retention periods (independent of the level of care), additional factors could include:

- criteria set as part of deposit compliance
- criteria set for initial curation
- Technical factors e.g. links to technical monitoring, format criteria, emulation approaches
- Semantic factors e.g. links to community monitoring, semantic artefacts, ontologies, controlled vocabularies etc.
- Preservation and ReAppraisal times (periods of time with start dates) and triggers (e.g. risk)

With the above feedback in mind, this draft working paper explores metadata characteristics of repositories and digital objects that would support the curation and preservation levels. A reference table in Appendix A. supports prose descriptions in the main body of the text.

---

[3] L'Hours, H., & Bell, D. (2023). UK Data Service (UKDS) Response to the CoreTrustSeal Curation & Preservation Levels Discussion Paper (v01.00). Zenodo. https://doi.org/10.5281/zenodo.7828046

Implied Metadata Characteristics

Transparency, and the development of ‘minimum viable’ or ‘ideal’ levels of practice, are challenging without agreement and implementation of standardised structured metadata. These include information on logical-technical (e.g. formats) and semantic (e.g. controlled vocabularies) criteria.

Levels of Repository & Object Care

The Curation & Preservation Levels imply explicit expressions of:

- Repository: Levels of Care Provided
- Digital Object: Level of Care Received

However, in many areas of digital object management practice expectations are implied or stated, but no specific metadata expression exists, or existing standards are not widely adopted. For example, all levels of care assume confidence in underlying storage, but there are no common metadata criteria to describe the number of copies, locations, media or integrity measures or the frequency of back up at either repository or object level.

Retention Information

All levels of care depend on the retention of the underlying digital objects. Repositories should declare high-level information about their approach to retention including any minimum or maximum retention periods and any exceptions including any trigger points (see ReAppraisal) that would impact those periods. At the object level, start times of retention periods should be included.

Appraisal/ ReAppraisal

At the point of initial appraisal, an object may be rejected, accepted as-is, or classified as requiring initial curation to meet desirable criteria. The criteria may include those necessary to support active preservation.

Not all digital objects retain their value over time and periodic reappraisal may be undertaken that changes the level of retention or preservation. Internal (change to repository mission/scope) or external (e.g. risk to a common file format) factors may trigger out-of-phase re-appraisal and these should be documented, including any impact on retention periods (see above). Changes to the level of retention or care should be transparent at the repository and object levels.

In addition to reappraisal periods, start date and triggers, all potential outcomes of a reappraisal should be defined, including changes to levels of care or decisions to delete or transfer a digital object.

D. Deposit Compliance

The deposit process acts as a 'gatekeeper' to the repository. Compliance checks against defined criteria provide a 'snapshot' of the state of the digital object before it is rejected, or accepted and custody transferred to the repository. Metadata should include Information about the criteria (including logical-technical and semantic) that an object should meet at the point of deposit (e.g. "sufficient metadata to meet required DataCite fields when assigning a DOI") and any exceptions to the compliance criteria (e.g. accepting a unique and high value deposit despite low quality).

C. Initial Curation

Metadata should include information about the criteria (including logical-technical and semantic) that the digital object must meet after curation. E.g. a specific level of FAIRness[4]. Any exceptions should be documented.

A. Active Preservation

In addition to a statement that the digital object is being cared for at level A, metadata should clarify the start date and minimum period the object will be cared for at this level. This should be supported by information about the community served and how engagement with that community is monitored.

Active Preservation periods may be shorter than retention periods. Any exceptions or triggers that would impact the preservation period should be documented.

For logical active preservation, information should be provided on the technical monitoring in place e.g. a link to repository documentation on file formats. For semantic preservation for understanding and reuse, information should be provided on the semantic artefacts (e.g. ontologies) that are in scope for the digital objects.

Conclusion

The levels of curation and preservation, and the implied associated metadata provide a simple sequential lifecycle perspective on digital objects and the repositories that care for them.

Most metadata examples above are alphanumeric values (e.g. dates, times) or selected from semantic artefacts such as controlled vocabularies (e.g. levels of care). These are likely to be machine-actionable for evaluation and assessment. Transparency of less-structured artefacts including policies and procedure is also required. Standardising and implementing the proposed metadata and supporting artefacts would strengthen links between repository and object information, and support transparent descriptions of care and reappraisal over time.

---

[4] 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

Appendix A: Table of Implied Object & Repository Metadata

NB: The type of information and local practice may influence whether information is provided in prose or in structured form and therefore whether it is machine actionable.

All of the below imply the need for rights information agreed between the repository, depositors and users about retention, deposit, curation and preservation.

Aspect	Implied Repository Level Information	Implied Object Level Information	Notes
Retention	Standard/Minimum/Maximum Retention Periods (Time)	Minimum Retention Period: Time Maximum Retention Period: Time Retention Start Date: YYYY-MM-DD	Essential information, independent of the level of care
	Exceptions to retention periods applied by the repository.	Exception to retention periods for this specific object.	Documented exceptions that might impact the retention period.
Level of Care	All Levels of Care provided by the repository to objects it holds (Z, D, C, A)	Current Level of Care received by this specific object (Z, D, C, A)	Levels of care from the controlled vocabulary of levels.
D. Deposit Compliance	Documented criteria that a digital object should meet at the point of deposit. Including Semantic (metadata, documentation, rights), technical (format), quality (formal data quality, scientific quality, ethical)	Link to repository level criteria in place when the object was deposited.	E.g. sufficient metadata to meet DataCite criteria when assigning a DOI. e.g., Metadata schema, version: URI e.g. Accepted formats, version: URI e.g. Quality criteria, version: URI
	Outcomes. All potential outcomes of the Deposit Compliance assessment.	Information about the outcome for this specific object.	Non-compliance causes a digital object to be rejected, or a non-compliant object may be moved forward to Initial Curation.

	Exceptions. All possible exceptions to the deposit compliance criteria and outcomes.	Information about any exceptions applied to this specific object.	Documented exceptions that might override the deposit compliance criteria. E.g. an object is accepted despite low quality as it is unique or of high value.
C. Initial Curation	Documented criteria that a digital object should meet at the end of the initial curation process. Including Semantic (metadata, documentation, rights), technical (format), quality (formal data quality, scientific quality, ethical)	Link to repository level criteria in place when the object was initially curated	May include the same criteria as set for deposit compliance or go beyond these.
	Exceptions. All possible exceptions to the criteria and outcomes	Information about any exceptions applied to this specific object.	
A. Active Preservation	Standard/Minimum/Maximum Preservation Periods (Time)	Preservation Period (Time) and Start (Date) for this specific object.	Active Preservation periods may be shorter than retention periods.
	Exceptions to Preservation periods applied by the repository.	Exception to preservation periods for this specific object.	Documented exceptions that might impact the preservation period.
	Preservation Action triggers applied by the repository	History of preservation actions applied to the specific digital object as part of provenance information.	Factors, including Semantic (metadata, documentation, rights), technical (format), quality (formal data quality, scientific quality, ethical) that might trigger a decision to take a preservation action on a digital object.

ReAppraisal	Standard/Minimum/Maximum ReAppraisal Periods (Time)	ReAppraisal Period (Time) and Start (Date) for this specific object.	Not all digital objects retain their value over time and periodic reappraisal may be undertaken that changes the level of retention or preservation. Include
	ReAppraisal Criteria	Link to repository level criteria in place when the object was ReAppraised	
	Outcomes. All potential outcomes of a ReAppraisal	History of reAppraisal outcomes for this specific digital object as part of provenance information.	ReAppraisal interacts with retention periods, preservation periods and preservation actions. *No change *Change (from curation/preservation level X to Y) *Decision not to retain: -Deletion -Transfer to an alternate repository
	Exceptions to ReAppraisal periods applied by the repository.	Exception to ReAppraisal periods for this specific object.	Documented exceptions that might impact the preservation period.
Community Served	Designated Community(s) served by the repository	Designated Community for the specific digital object	
Logical Technical	Technical Factors (formats)	Technical factors for this specific object	E.g. link to repository documentation on file formats. Including preservation action triggers such as file format risk where applicable

	How the technical landscape is monitored	Link to repository technical landscape monitoring information	
	How the technical needs of the user community are monitored	Link to repository community monitoring information	
Semantic	Semantic Factors (metadata, documentation, rights)	Semantic factors for this specific object	The semantic artefacts (e.g. ontologies) that are in scope for the digital objects. Including preservation action triggers such as deprecated, new or updated ontologies where applicable
	How the semantic landscape is monitored	Link to repository semantic landscape monitoring information	
	How the semantic needs of the user community are monitored	Link to repository community monitoring information	
Quality Factors	Quality Factors (formal data quality, scientific quality, ethical)	Quality factors for this specific object	
	How the quality landscape is monitored	Link to repository quality landscape monitoring information	
	How the quality needs of the user community are monitored	Link to repository community monitoring information	

Appendix B: CoreTrustSeal Curation & Preservation Level Descriptions in Full

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.

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.[5] Digital objects that do not meet these criteria may be rejected, or moved forward 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.

[5] 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.

Authors & Acknowledgements

The authors would like to acknowledge the foundational work provided by the CoreTrustSeal Standards and Certification Board[6] on the Levels of Curation and Preservation.

Author & Identifiers	Affiliations and Associations
Hervé L'Hours https://orcid.org/0000-0001-5137-3032	UK Data Archive[7], UK Data Service[8], University of Essex[9].
Mari Kleemola https://orcid.org/0000-0001-8855-5075	Finnish Social Science Data Archive (FSD[10]), Tampere University[11]
Jonas Recker https://orcid.org/0000-0001-9562-3339	GESIS[12] – Leibniz Institute for the Social Sciences

[6] https://www.coretrustseal.org/about/standards-and-certification-board/
[7] https://www.data-archive.ac.uk/
[8] https://ukdataservice.ac.uk/
[9] https://www.essex.ac.uk/about/faculty-of-social-sciences
[10] https://www.fsd.tuni.fi/en/
[11] https://www.tuni.fi/en
[12] https://www.gesis.org/