CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022

Background & General Guidance

The *CoreTrustSeal Trustworthy Data Repositories Requirements* describe the characteristics of trustworthy repositories. All Requirements are mandatory and evaluated as standalone items. Although some overlap is unavoidable, duplication of evidence sought for each Requirement has been kept to a minimum. The options in checklists (e.g., repository type and curation level) are not considered to be comprehensive and may be refined in the future. Applicants are encouraged to add 'other' options.

Each Requirement is accompanied by Guidance text describing the information and evidence that applicants must provide to enable an objective review.

The applicant must indicate a compliance level for each of the Requirements:

0 – Not applicable
1 – The repository has not considered this yet
2 – The repository has a theoretical concept
3 – The repository is in the implementation phase
4 – The guideline has been fully implemented in the repository

Compliance levels are an indicator of the applicant's self-assessed progress, but reviewers judge compliance against response statements and supporting evidence. If an applicant believes a Requirement is not applicable (0), then this must be justified in detail. Compliance Levels of 1 or 2 are not sufficient for a successful application. Certification may be granted if some Requirements are in the implementation phase (3).

Responses statements provided by applicants should include links to supporting evidence online. As the core certification process does not include a site visit by an auditor, such publically available evidence provides transparent assurance of good practice. URL links should be verified immediately before submitting applications.

All responses must be in English. Although attempts will be made to match reviewers to applicants in terms of language and discipline, this is not always possible. Full translations of evidence are not required, but if non-English evidence is provided, then an English summary must be included in the response statement.

No sensitive information disclosure is required to acquire the CoreTrustSeal, but provisions are made within the certification process for repositories that want to share evidence materials also containing confidential information.

The CoreTrustSeal is valid for three years from the date it is awarded. Though repository systems and capabilities evolve continuously according to technology and user needs, they might not undergo major changes in this timeframe. An organization with well-managed business processes and records should be able to reapply with minimal revisions after three years unless:

- The organization, its data collection, or Designated Community has changed significantly.
- The CoreTrustSeal Requirements have been updated in ways that impact the applicant.

The CoreTrustSeal Requirements are subject to review and revision every three years. This does not affect a successful applicant until they seek renewal.

Glossary of Terms

Please refer to the CoreTrustSeal Trustworthy Data Repositories Requirements Glossary: https://doi.org/10.5281/zenodo.3632563.

Requirements

Background Information

Context

R0. Please provide context for your repository.

– Repository Type. Select all relevant types from:

- **Domain or subject-based repository**
- **Institutional repository**
- **National repository system, including governmental**
- **Publication repository**
- **Library**
- **Museum**
- **Archive**
- **Research project repository**
- **Other (Please describe)**

– Brief Description of Repository

– Brief Description of the Designated Community

– 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**

Comments

– Insource/Outsource Partners. If applicable, please list them.

– Summary of Significant Changes Since Last Application (if applicable)

– Other Relevant Information

Response

Guidance:

The information in this section provides the background and context needed by reviewers to fully assess the responses to the other Requirements. It is therefore of vital importance to the entire application that detailed responses are given to each question. Please select from among the options and provide details for the items that appear in the Context Requirement.

(1) Repository Type. This item will help reviewers understand what function your repository performs. Choose the best match for your repository type (select all that apply). If none of the categories is appropriate, feel free to provide another descriptive type. You may also provide further details to help the reviewer understand your repository type.

(2) Brief Description of Repository. Provide a short overview of the repository; in particular, please add information on the type of data accepted by the repository (i.e., the scope of its collection). If the repository has outsource partners, is part of a network, or of a parent organization, the response should ideally include a diagram and description of the overarching organizational structure.

(3) Designated Community. A clear definition of the Designated Community demonstrates that the applicant understands the scope, knowledge base, and methodologies—including preferred software/formats—of the user community or communities they are targeting. Please make sure that the response is sufficiently specific to enable reviewers to assess the adequacy of the curation and preservation measures described throughout the application.

(4) Level of Curation. This item is intended to elicit whether the repository distributes its content to data consumers without any changes, or whether the repository adds value by enhancing the content in some way. All levels of curation assume initial deposits are retained unchanged and that edits are only made on copies of those originals. Annotations/edits must fall within the terms of the license agreed with the data producer and be clearly within the skillset of those undertaking the curation. Thus, the repository will be expected to demonstrate that any such annotations/edits are undertaken and documented by appropriate experts and that the integrity of all original copies is maintained. Knowing this will help reviewers in assessing other certification Requirements. Further details can be added that would help to understand the levels of curation you undertake.

(5) Insource/Outsource Partners. Please provide a list of Partners that your organization works with, describing the nature of the relationship (organizational, contractual, etc.), and whether the Partner has undertaken any trustworthy repository assessment. If a function or supporting evidence is not under the direct control of the applicant then it falls into this category. This may be with a host organization or other 'insourcing' relationship, or through outsourcing or other dependency on a third-party. Such relationships may include, but are not limited to: any services provided by an institution you are part of, storage provided by others as part of multicopy redundancy, or membership in organizations that may undertake stewardship of your data collection when a business continuity issue arises. Moreover, please list the certification requirements for which the Partner provides all, or part of, the relevant functionality/service, including any contracts or Service Level Agreements in place. Because outsourcing will almost always be partial, you will still need to provide appropriate evidence for certification Requirements that are not outsourced and for the parts of the data lifecycle that you control. Qualifications/certifications—including, but not limited to, the CoreTrustSeal certification (and its predecessors)—are preferred for outsource partners. However, it is not a necessity for them to be certified. We understand that this can be a complex area to define and describe, but such details are essential to ensure a comprehensive review process.

(6) Summary of Significant Changes Since Last Application. CoreTrustSeal certification has an expectation of continuous improvement. Repositories undergoing recertification should highlight briefly to the reviewers any significant changes in technical systems, Designated Community, funding, and so on during the previous three years. In doing so, please refer to any comments given to you by the reviewers of your previous CoreTrustSeal application. Detailed information on a change should be added to the appropriate Requirement.

(7) Other Relevant Information. The repository may wish to add extra contextual information that is not covered in the Requirements but that may be helpful to the reviewers in making their assessment. For example, you might describe:

- The usage and impact of the repository data holdings (citations, use by other projects, etc.).
- A national, regional, or global role that the repository serves.
- Any global cluster or network organization that the repository belongs to.

Organizational Infrastructure

1. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

Response

Guidance:

Repositories take responsibility for stewardship of digital objects, and for ensuring that materials are held in the appropriate environment for appropriate periods of time. Depositors and users must be clear that preservation of and continued access to the data is an explicit role of the repository.

For this Requirement, please describe:

- Your organization’s mission in preserving and providing access to data, and include links to explicit statements of this mission.
- The level of approval that the mission has received within the organization.

Evidence for this Requirement could take the form of an approved public mission statement, roles mandated by funders, policy statement signed off by governing board.

2. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

Response

Guidance:

Repositories must have an appropriate rights model covering data access and use, communicate about them with users, and monitor compliance. This Requirement relates to the access regulations and applicable licenses set by the data repository itself, as well as any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information. Evidence should demonstrate that the repository has sufficient controls in place according to the access criteria of their data holdings, as well as evidence that any relevant licenses or processes are well managed.

For this Requirement, please describe:

- License agreements in use.
- Conditions of use (Intellectual Property Rights, distribution, intended use, protection of sensitive data, etc.).
- Documentation on measures in the case of noncompliance with conditions of access and use.

Note that if all data holdings are completely public and without conditions imposed on users—such as attribution requirements or agreement to make secondary analysis openly available—then it can simply be stated.

The ethical and privacy provisions that impact on licenses are dealt with in R4 (Confidentiality/Ethics). Assurance that deposit licenses provide sufficient rights for the repository to maintain, preserve, and offer access to data should be covered under R10 (Preservation Plan).

3. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

Response

Guidance:

This Requirement covers the governance related to continued operation of the repository over time and during disasters, as well as evidence in relation to succession planning; namely, the measures in place to ensure access to and availability of data holdings, both currently and in the future. Reviewers are seeking evidence that preparations are in place to address the risks inherent in changing circumstances, including in mission and/or scope.

For this Requirement, please describe:

- The level of responsibility undertaken for data holdings, including any guaranteed preservation periods.
- The medium-term (three- to five-year) and long-term (> five years) plans in place to ensure the continued availability and accessibility of the data. In particular, both the response to rapid changes of circumstance and long-term planning should be described, indicating options for relocation or transition of the activity to another body or return of the data holdings to their owners (i.e., data producers). For example, what will happen in the case of cessation of funding, which could be through an unexpected withdrawal of funding, a planned ending of funding for a time-limited project repository, or a shift of host institution interests?

Evidence for this Requirement should relate specifically to governance. The technical aspects of business continuity, and disaster and succession planning should be covered in R15 (Technical infrastructure).

4. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

Response

Guidance:

Adherence to ethical norms is critical to responsible science. Disclosure risk—for example, the risk that an individual who participated in a survey can be identified or that the precise location of an endangered species can be pinpointed—is a concern that many repositories must address. Evidence should demonstrate that the repository has good practices for data with disclosure risks, including guidance for depositors and users. This is necessary to maintain the trust of those agreeing to have personal/sensitive data stored in the repository.

For this Requirement, responses should include evidence related to the following questions:

- How does the repository comply with applicable disciplinary norms?
- Does the repository request confirmation that data collection or creation was carried out in accordance with legal and ethical criteria prevailing in the data producer's geographical location or discipline (e.g., Ethical Review Committee/Institutional Review Board or Data Protection legislation)?
- Are special procedures applied to manage data with disclosure risk?
- Are data with disclosure risk managed appropriately to limit access?
- Are data with disclosure risk distributed under appropriate conditions?
- Are procedures in place to review disclosure risk in data, and to take the necessary steps to either anonymize files or to provide access in a secure way?
- Are staff trained in the management of data with disclosure risk?
- Are there measures in place if conditions are not complied with?
- Does the repository provide guidance in the responsible deposit, download, and use of disclosive, or potentially disclosive data?

This Requirement is about the ethical and privacy provisions that impact the creation, curation, and use of the data. Details on any licenses in alignment with such ethical and privacy provisions should be covered in R2 (Licenses).

5. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

Response

Guidance:

Repositories need funding to carry out their responsibilities, along with a competent staff who have expertise in data archiving. However, it is also understood that continuity of funding is seldom guaranteed, and this must be balanced with the need for stability.

For this Requirement, responses should include evidence related to the following:

- The repository is hosted by a recognized institution (ensuring long-term stability and sustainability) appropriate to its Designated Community.
- The repository has sufficient funding, including staff resources, IT resources, and a budget for attending meetings when necessary. Ideally this should be for a three- to five-year period.
- The repository ensures that its staff have access to ongoing training and professional development.
- The range and depth of expertise of both the organization and its staff, including any relevant affiliations (e.g., national or international bodies), is appropriate to the mission.

Full descriptions of the tasks performed by the repository—and the skills necessary to perform them—may be provided, if available. Such descriptions are not mandatory, however, as this level of detail is beyond the scope of core certification.

Access to objective expert advice beyond that provided by skilled staff is covered in R6 (Expert guidance).

6. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Compliance Level:

Response

Guidance:

An effective repository strives to accommodate evolutions in data types, data volumes, and data rates, as well as to adopt the most effective new technologies in order to remain valuable to its Designated Community. Given the rapid pace of change, it is therefore advisable for a repository to secure the advice and feedback of expert users on a regular basis to ensure its continued relevance and improvement.

For this Requirement, responses should include evidence related to the following questions:

- Does the repository have in-house advisers, or an external advisory committee that might be populated with technical, curation, data science, and disciplinary experts?
- How does the repository communicate with the experts for advice?
- How does the repository communicate with its Designated Community for feedback?

This Requirement seeks to confirm that the repository has access to objective expert advice beyond that provided by skilled staff mentioned in R5 (Organizational infrastructure).

Digital Object Management

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

Response

Guidance:

The repository should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access. This Requirement covers the entire data lifecycle within the repository.

To protect the integrity of data and metadata, any intentional changes to data and metadata should be documented, including the rationale and originator of the change. Measures should be in place to ensure that unintentional or unauthorized changes can be detected and correct versions of data and metadata recovered.

Authenticity covers the degree of reliability of the original deposited data and its provenance, including the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

For this Requirement, responses on data integrity should include evidence related to the following:

- Description of checks to verify that a digital object has not been altered or corrupted (i.e., fixity checks) from deposit to use.
- Documentation of the completeness of the data and metadata.
- Details of how all changes to the data and metadata are logged.
- Description of version control strategy.
- Usage of appropriate international standards and conventions (which should be specified).

Evidence of authenticity management should relate to the following questions:

- Does the repository have a strategy for data changes? Are data producers made aware of this strategy?
- Does the repository maintain provenance data and related audit trails?
- Does the repository maintain links to metadata and to other datasets? If so, how?
- Does the repository compare the essential properties of different versions of the same file? How?
- Does the repository check the identities of depositors?

8. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

Response

Guidance:

The appraisal function is critical to evaluate whether data meet all criteria for selection and to ensure appropriate management for their preservation. Appraisal and reappraisal over time ensure data remain relevant and understandable to the Designated Community.

For this Requirement, responses should include evidence related to the following questions:

- Does the repository use a collection development policy to guide the selection of data for archiving?
- What approach is used for data that do not fall within the mission/collection profile?
- Does the repository have procedures in place to determine that the metadata required to interpret and use the data are provided?
- Is there any automated assessment of metadata adherence to relevant schemas?
- What is the repository’s approach if the metadata provided are insufficient for long-term preservation?
- Does the repository publish a list of preferred formats?
- Are checks in place to ensure that data producers adhere to the preferred formats?
- What is the approach towards data that are deposited in non-preferred formats?
- What is the process for removing items from your collection, also keeping in mind impact on existing persistent identifiers?

This Requirement covers the selection criteria applied at the point of deposit. Data quality and improvement during the curation process should be covered under R11 (Data quality).

9. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

Response

Guidance:

Repositories need to store data and metadata from the point of deposit, through the ingest process, to the point of access. Repositories that perform digital preservation must offer ‘archival storage’ in OAIS terms.

For this Requirement, responses should include evidence related to the following questions:

- How are relevant processes and procedures documented and managed?
- Does the repository have a clear understanding of all storage locations and how they are managed?
- Does the repository have a strategy for multiple copies? If so, what is it?
- Are risk management techniques used to inform the strategy?
- What checks are in place to ensure consistency across archival copies?
- How is deterioration of storage media handled and monitored?

Details on the technical implementation of storage should be covered in R15 (Technical infrastructure), and specific arrangements for physical and logical security in R16 (Security).

10. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

Response

Guidance:

The repository, data depositors, and Designated Community need to understand the level of responsibility undertaken for each deposited item in the repository. The repository must have the rights to undertake these responsibilities. Procedures must be documented and their completion assured.

For this Requirement, responses should include evidence related to the following questions:

- Does the repository have a documented approach to preservation?
- Is the level of responsibility for the preservation of each item understood? How is this defined?
- Are plans related to future migrations or similar measures to address the threat of obsolescence in place?
- Does the contract between depositor and repository provide for all actions necessary to meet the responsibilities?
- Is the transfer of custody and responsibility handover clear to the depositor and repository?
- Does the repository have the rights to copy, transform, and store the items, as well as provide access to them?
- Are actions relevant to preservation specified in documentation, including custody transfer, submission information standards, and archival information standards?
- Are there measures to ensure these actions are taken?

Rights concerning data access and use, and the monitoring of their compliance should be covered under R2 (Licenses).

11. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

Response

Guidance:

Repositories must ensure there is sufficient information about the data for the Designated Community to assess the quality of the data. Quality assessment becomes increasingly relevant when the Designated Community is multidisciplinary, where users may not have the personal experience to make an evaluation of quality from the data alone. Repositories must be able to evaluate completeness and quality of the data and metadata.

Data, or associated metadata, may have quality issues relevant to their research value, but this does not preclude their use if a user can make a well-informed decision on their suitability through provided documentation.

For this Requirement, please describe:

- The approach to data and metadata quality taken by the repository.
- Does the repository have quality control checks to ensure the completeness and understandability of data deposited? If so, please provide references to quality control standards and reporting mechanisms accepted by the relevant community of practice, and include details of how any issues are resolved (e.g., are the data returned to the data provider for rectification, fixed by the repository, noted by quality flags in the data file, and/or included in the accompanying metadata?).
- The ability of the Designated Community to comment on, and/or rate data and metadata.
- Whether citations to related works or links to citation indices are provided.

This Requirement refers to data quality standards and assurance during curation. Selection criteria are covered in R8 (Appraisal).

12. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

Response

Guidance:

To ensure the consistency of practices across datasets and services and to avoid ad hoc actions, workflows should be defined according to the repository's activities and clearly documented. Provisions for managed change should be in place. The OAIS reference model can help to specify the workflow functions of a repository.

For this Requirement, responses should include evidence related to the following:

- Workflows/business process descriptions.
- Clear communication to depositors and users about handling of data.
- Levels of security and impact on workflows (guarding privacy of subjects, etc.).
- Qualitative and quantitative checking of outputs.
- The types of data managed and any impact on workflow.
- Decision handling within the workflows (e.g., archival data transformation).
- Change management of workflows.

This Requirement confirms that all workflows are documented.

13. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

Response

Guidance:

Effective data discovery is key to data sharing. Once discovered, datasets should be referenceable through full citations, including persistent identifiers to help ensure that data can be accessed into the future.

For this Requirement, responses should include evidence related to the following questions:

- Does the repository offer search facilities?
- Does the repository maintain a searchable metadata catalogue to appropriate (internationally agreed) standards?
- What persistent identifier systems does the repository use?
- Does the repository facilitate machine harvesting of the metadata?
- Is the repository included in one or more disciplinary or generic registries of resources?
- Does the repository offer recommended data citations?

14. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

Response

Guidance:

Repositories must ensure that data continues to be understood and used effectively into the future despite changes in technology and the Designated Community's knowledge base. This Requirement evaluates the measures taken to ensure that data are reusable.

For this Requirement, responses should include evidence related to the following questions:

- Which metadata are provided by the repository when the data are accessed?
- How does the repository ensure continued understandability of the data?
- Are data provided in formats used by the Designated Community? Which formats?
- Are measures taken to account for the possible evolution of formats?

The concept of 'reuse' is critical in environments in which secondary analysis outputs are redeposited into a repository alongside primary data, since the provenance chain and associated rights issues may then become increasingly complicated.

Technology

15. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

Response

Guidance:

Repositories need to operate on reliable and stable core infrastructures that maximizes service availability. Furthermore, hardware and software used must be relevant and appropriate to the Designated Community and to the functions that a repository fulfils. The OAIS reference model specifies the functions of a repository in meeting user needs.

For this Requirement, responses should include evidence related to the following questions:

- What standards does the repository use for reference? Are these international and/or community standards? How often are these reviewed?
- How are the standards implemented? Are there any significant deviations from the standard? If so, please explain.
- Does the repository have a plan for infrastructure development? If so, what is it?
- Is a software inventory maintained and is system documentation available?
- Is community-supported software in use? Please describe.
- Are availability, bandwidth, and connectivity sufficient to meet the needs of the Designated Community?
- Does the repository have a disaster plan and a business continuity plan? In particular, are procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage? What are they?

The governance aspects of business continuity, disaster planning, and succession planning should be covered in R3 (Continuity of access). Details on the storage process should be covered in R9 (Documented storage procedures). Security arrangements are covered in R16 (Security).

16. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

Response

Guidance:

The repository should analyze potential threats, assess risks, and create a consistent security system. It should describe damage scenarios based on malicious actions, human error, or technical failure that pose a threat to the repository and its data, products, services, and users. It should measure the likelihood and impact of such scenarios, decide which risk levels are acceptable, and determine which measures should be taken to counter the threats to the repository and its Designated Community. This should be an ongoing process.

For this Requirement, please describe:

- Your IT security system, employees with roles related to security (e.g., security officers), and any risk analysis tools (e.g., DRAMBORA[1]) you use.
- What levels of security are required, and how these are supported.
- Any authentication and authorization procedures employed to securely manage access to systems in use (e.g., Shibboleth, OpenAthens).

The storage processes and technical infrastructure that utilize these security measures should be covered in R9 (Documented storage procedures) and R15 (Technical Infrastructure), respectively.

[1] https://www.repositoryaudit.eu/

Applicant Feedback

Comments/feedback

These Requirements are not seen as final, and we value your input to improve the CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response