Core Trustworthy Data Repositories Extended Guidance

- Core Trustworthy Data Repositories Requirements for 2017–2019
- Extended Guidance

June 2018

Core Trustworthy Data Repositories Extended Guidance

Version 1.1: June 2018

Table of Contents

Introduction

This document consists of the **Core Trustworthy Data Repositories Requirements** for 2017–2019 with introductory paragraphs on Background & General Guidance, which are set by the CoreTrustSeal Board and remain unchanged for the period 2017–2019. The fixed text is recognizable by the boxes drawn around it.

The document furthermore contains the **Extended Guidance** for CoreTrustSeal reviewers and applicants. This information will be subject to change and updates during the period for which the seal has been awarded, according to the needs of the community and approval by the CoreTrustSeal Board. The document also contains a reference to the Glossary of Terms.

Background & General Guidance

The *Core Trustworthy Data Repositories Requirements* are intended to describe the characteristics of trustworthy repositories. As such, all Requirements are mandatory and are equally weighted, standalone items. Although some overlap is unavoidable, duplication of evidence sought for each Requirement has been kept to a minimum where possible. The choices contained in checklists (e.g., repository type and curation level) are not considered to be comprehensive, and additional space is provided in all cases for the applicant to add 'other' (missing) options. This and any comments given may then be used to refine such lists in the future.

Each Requirement in the Catalogue is accompanied by guidance text to assist applicants in providing sufficient evidence that their repositories meet the Requirement, outlining the types of information that a reviewer will expect in order to perform an objective assessment. Furthermore, 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 provide a useful part of the self-assessment process, but all applicants will be judged against statements supported by appropriate evidence and not against self-assessed compliance levels. In this regard, if the applicant believes a Requirement is not applicable, the reason for this must be documented in detail. Note also that compliance levels 1 and 2 can be valid for internal self-assessments, while certification may be granted if some guidelines are considered to be at level 3—in the implementation phase—since the Requirements include an assumption of a repository's continuous improvement.

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. If evidence is in another language, an English summary must be provided in the self-assessment.

Because core certification does not involve a site visit, the Requirements should be supported by links to public evidence. Nevertheless, it is understood that for reasons such as security, it may not always possible to include all information on an organization's website, and provisions are made within the certification process for repositories that want sensitive parts of their evidence to remain confidential.

Repositories are required to be reassessed every three years. It is recognized that while basic systems and capabilities evolve continuously according to technology and user needs, they

might not undergo major changes in this timeframe. However, the Trustworthy Repository ISO standard (ISO 16363) has a five-year review cycle, and a shorter period is considered necessary for a core trust standard to allow for possible modifications and corrections. Hence, an organization with well-managed records and business processes should reasonably expect to be able to submit an application with only minimal revisions after three years, unless the Requirements themselves have been updated within the intervening period.

Glossary of Terms

Please refer to the Core Trustworthy Data Repositories Requirements Glossary: https://goo.gl/rQK5RN.

General Extended Guidance

These guidelines are aimed at the development of common standards and procedures for reviewing the self-assessments for the CoreTrustSeal certification of repositories. They are primarily intended for giving reviewers guidance when reviewing the self-assessments, but are also made available for everyone to consult as they might be helpful for applicants when formulating a self-assessment.

Introduction: General Points

Effort for a first round application should be supported by ongoing maintenance of publicly accessible business information. Revising all of your business information during the assessment is not what is being looked for. The (ongoing) management of information necessary to run your services should be sufficient to maintain certification.

Concerning the level of compliance, the reviewers may decide the level of compliance is different from the one selected for a Requirement. If the reviewers change the selected level to a *lower value,* the reason will be explained to and agreed upon with the applicant. There is then an expectation of progress for that Requirement when the certification is renewed.

Guidance should always be indicative. Reviewers will check that each item of guidance is addressed in the evidence statement. As it is not possible to cover every scenario in the general guidance, each applicant is expected to extend their response to provide local conditions and context. Reviewers are looking for clear, open statements of evidence specific to the applicant. Not necessarily all bullet points in all Requirements are mandatory; final judgment depends on the completeness and quality of the answer in the self-assessment of a specific Requirement.

Missing information/evidence

If information is missing to such a degree that it is impossible to judge whether the specific Requirement is met, then this Requirement should be sent back to the applicant with an explanation. No compliance level will be given at this stage by the reviewer.

This follows from the paragraph Background & General Guidance section, where it is stated:

‘Compliance levels provide a useful part of the self-assessment process, but all applicants will be judged against statements supported by appropriate evidence; not against self-assessed compliance levels’.

This means that specific assertions should be supported by links to public evidence as much as possible. In other words, evidence is essential. It is problematical and/or unacceptable when information is (almost) completely missing, not sufficiently or clearly described, or is mentioned or referred to under another Requirement. Any deviations from this should be commented on explicitly.

It also means that possible familiarity with or inside knowledge of the repository by the reviewer cannot play a role when judging the available evidence.

Understandability of documentation

Reviewers and readers of the final review, which is made public on both the repository and the CoreTrustSeal websites, should be able to understand the procedure without detailed reading of supportive evidence. For lengthy documents presented as evidence, the applicant should refer to the specific sections that provide the supporting evidence and quote/summarize the information in the application, as appropriate. A clear and consistent description about the organizational approach as a whole is generally helpful.

Non-English language documentation

For linked evidence in languages other than English, more detail must be provided in the evidence statements of the self-assessment. Documentation in languages other than English is acceptable if its content is sufficiently and clearly explained in an English summary. This summary can be quite brief for certain types of documents (e.g., a list of preferred formats), but should be longer for others (e.g., a Preservation Policy document).

This follows directly from: ‘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. If evidence is in another language, an English summary must be provided in the self-assessment.’

Confidentiality of internal documents

Information that is confidential, commercially sensitive, or poses a security risk, cannot be considered as the kind of public evidence required to meet the Core Trustworthy Data Repositories Requirements. This also applies to documents that are only available on the intranet of the repository. Applicants may have evidence documents containing such information alongside relevant evidence, and these should be submitted confidentially to the reviewers and the documents named and described in the application[1]. Applicants should aim at revising future documentation to remove confidential information so that it can be made public and considered during the next review.

If documentation is for internal use only, or if it does not exist or is in progress, a date of public availability should be stated. This documentation will be judged in the same way as public information. No explicit reference to these internal documents may be made in the comments by the reviewers, who will accept this on the overall understanding that the documentation will be made public within the promised schedule. All deadlines will be part of the overall reviewer comments and will be checked once a renewal is submitted. If again in the next round, the information is or cannot be made public by the applicant, they should be asked to publish at least a summary of the document.

[1] To safeguard the confidentiality of reviewers, confidential documents should always be submitted via the CoreTrustSeal e-mail address.

Application structure and length

The audience of CoreTrustSeal applications (i.e., self-assessments) is initially CoreTrustSeal reviewers and the Board, but successful applications are made publicly available. Applicants should therefore keep all of these audiences in mind. Applications should not respond to each item of guidance in a question-and-answer format. Applications should include prose responses to each Requirement, incorporating relevant elements of the Guidance and Extended Guidance provided.

The CoreTrustSeal Board understands that applicants of the CoreTrustSeal come from a wide range of organizations of varying mission, size, and complexity (in terms of both organizational structure and data collection variety), and so there may be relevant topics and evidence required that is not covered by the Guidance. We also understand that space is needed to explain the relevance of evidence provided; especially, if not available in English. The Board does not set minimum or maximum lengths for responses, but in its experience, even the most complex evidence statements are at the lower end of the 500–800 word range. Wherever possible, evidence statements should be supported by public links to the documentation used to govern your organization and manage your digital objects. It is this public evidence that offers the most assurance an organization manages its collections as a Trustworthy Data Repository.

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/Archives**
- **Research project repository**
- **Other (Please describe)**

Brief Description of Repository

– *Brief Description of the Repository's 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

– *Outsource Partners*. If applicable, please list them.

– *Other Relevant Information*

Response

Guidance:
To assess a repository, reviewers need some information about the repository to set it in context. 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) Repository's Designated Community. This item will be useful in assessing how the repository interacts and communicates with its target community. Please make sure that the response is specific—for example, 'quantitative social science researchers and instructors.'

(3) 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 licence 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.

(4) Outsource Partners. Please provide a list of Outsource 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. 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 DSA or WDS certifications—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.

(5) 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.

Extended Guidance R0.

Repository Type
Selection of more than one type should be supported by a description of the repository explaining how these multiple roles are fulfilled. Ideally, such a description would provide additional context by describing the data types, formats, disciplines, and so on managed by the repository.

Brief Description of the Repository's Designated Community
The glossary contains a formal definition of 'Designated Community'. However, the following description is also provided to help elucidate what the meaning of 'Designated Community' in this context:

A clear Designated Community demonstrates that the applicant understands the scope, knowledge base, and methodologies—including preferred software/formats—they are targeting. Ideally an understanding of what curation needs to be done (additional context, preferred formats etc.) to serve that Community is indicated throughout the evidence, alongside a demonstration that the applicant monitors and responds to changes in the needs of that Community.

Evidence across the Requirements should demonstrate interactions with—and responsiveness to—the Designated Community; especially, as regards to changes in its needs. A repository with a

very specific, narrow Designated Community might state the expected knowledge base (e.g., degree/level of understanding of statistics or genetics). A very broad designated community (e.g., the general public) would imply that the repository has a wider range of contextual documentation to ensure its data can be understood by everyone.

Level of Curation Performed
More than one option (A, B, C, or D) of the level (or extent) of curation can be picked, depending on the type of data and agreement with the depositor.
Reviewers will expect a higher level of formal provenance, integrity, and version management (change logs etc.) in the case of D.

Outsource Partners
If more than one partner is involved, a diagram to indicate the full scale of the outsourcing process is useful to assist the reviewers in their understanding. Having more than one outsource partner (e.g., one for storage, one for maintaining websites) is okay as long as all the relations are clearly indicated, preferably in a diagram. Reviewers will ask for this section to be revised if the evidence statements later in the self-assessment refer to entities not mentioned here.

Other Relevant Information
A repository might refer here to its re3data record (http://www.re3data.org/), number of staff, size of collection, average number of downloads, its evolution over time, business or funding model, and so on.

Organizational Infrastructure

I. 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 within the organization that such a mission statement has received (e.g., approved public statement, roles mandated by funders, policy statement signed off by governing board).

Extended Guidance R1.

If data management is not referred to in the mission statement, then, as a rule, this Requirement cannot have a compliance level of 3 or higher.

II. Licenses

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

Compliance Level:

Response

Guidance:
Repositories must maintain all applicable licenses 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. Reviewers will be seeking evidence that the repository has sufficient controls in place according to the access criteria of their data holdings, as well as evidence that any relevant licences or processes are well managed.

For this Requirement, please describe:

- License agreements in use.
- Conditions of use (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.

This Requirement must be read in conjunction with R4 (Confidentiality/Ethics) to the extent that ethical and privacy provisions impact on the licenses. Assurance that deposit licences provide sufficient rights for the repository to maintain, preserve, and offer access to data is covered under R10 (Preservation Plan).

Extended guidance R2.

Access and use conditions could be set differently: either as standard terms and conditions, or as differentiated for particular depositors or datasets. These could cover the level of curation, what is the liability level, the level of responsibility taken for the data, limitations on use, limits on usage environment (safe room, secure remote access), and limits on types of users (approved researcher, has received training, etc.). The Creative Commons licences (https://creativecommons.org/), including CC 0 Waiver and public domain data, could be used as a reference here, but other alternatives are also possible.

While it may be challenging to identify instances of noncompliance, some consideration should be given to the consequences if noncompliance is detected (e.g., sanctions on current or future access/use of data). In the case of sensitive personal data disclosure, there may be severe legal penalties that impact both the user and repository. Ideally, repositories should have a public policy in place for noncompliance.

The minimum compliance level should be 4, if the applicant is currently providing access to data.

III. 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 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.

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 more to governance than to the technical information that is needed in R10 (Preservation plan) and R14 (Data reuse), and should cover the situation in which R1 (Mission/Scope) changes. This Requirement contrasts with R15 (Technical infrastructure) and R16 (Security) in that it covers full business continuity of the preservation and access functions.

Extended Guidance R3.

The reviewer is looking for information to understand the level of responsibility for data. For example, are you the primary or only custodian? Is the depositor responsible as well? Does the repository promise to provide access, preservation, and/or data storage to some minimum quality level for some minimum time period? This information helps the reviewer to judge if the repository is sustainable in terms of its finances and processes; in particular, the continuity of its collections and responsibilities in the case of a major business continuity failure. The responsibility for sustainability may not lie in the hands of the repository itself, but a higher, overarching (or umbrella) organization. If so, this should be clearly indicated. Moreover, if the repository is part of such a larger (umbrella) organization, has this or any other organization (e.g., National Archives) guaranteed that it will take over the responsibility in the case of major business continuity failure? If there is no formal, written agreement between the respository and such an organization, then the compliance level can be at a maximum of 3 only.

IV. 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 sought is concerned with not only having good practices for data with disclosure risks, but also the necessity 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 stored 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 use of disclosive, or potentially disclosive data?

Evidence for this Requirement should be in alignment with provisions for the procedures stated in R12 (Workflows) and for any licenses in R2 (Licences).

Extended Guidance R4.

All organizations responsible for data have an ethical duty to manage them to the level expected by the scientific practice of its Designated Community. For repositories holding data about individuals, businesses, or other organizations, there are in addition expectations that the rights of the data subjects will be protected. These will be both of a legal and ethical nature.

Disclosure of these data could also present a risk of personal harm, a breach of commercial confidentiality, or the release of critical information (e.g., the location of protected species or an archaeological site).

Minimum compliance level should be a 4 if the repository is currently providing access to personal data.

Reviewers expect to see evidence that the applicant understands their legal environment and the relevant ethical practices, and that they have documented procedures.

V. 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.

Extended Guidance R5.

The description of this Requirement should contain evidence describing the organization's governance/management decision-making processes and the entities involved. Staff should have appropriate training in data management to ensure consistent quality standards. It is also important to know what proportion of staff is employed on a permanent or temporary basis and how this might affect the professional quality of the repository, particularly for long-term preservation.

To what degree is funding structural or project-based? Can this be expressed in FTE numbers?

How often does periodic renewal occur?

VI. 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 in the research data environment, 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 members, data science experts, 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).

Extended Guidance R6.

The reviewer is looking for evidence that the repository is linked to a wider network of expertise in order to demonstrate access to advice and guidance for both its day-to-day activities and the monitoring of potential new challenges on the horizon (science and technology watch). Part of this information may already have been given under ‘R0. Brief Description of the Repository’s Designated Community’ and ‘Other relevant information’. If so, then please refer to it.

Digital Object Management

VII. 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.

Integrity ensures that changes to data and metadata are documented and can be traced to the rationale and originator of the change.

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).
- 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 follow 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?

This Requirement covers the entire data lifecycle within the repository, and thus has relationships with workflow steps included in other requirements—for example, R8 (Appraisal) for ingest, R9 (Documented storage procedures) and R10 (Preservation plan) for archival storage, and R12–R14 (Workflows, Data discovery and identification, and Data reuse) for dissemination. However, maintaining data integrity and authenticity can also be considered a mindset, and the responsibility of everyone within the repository.

Extended Guidance R7.

A clear and complete context section is important for all Requirements but this is especially the case for R7. The organization of the curation and the types of data will help guide the reviewer

expectation. The reviewer will benefit from a clear overview of the processes and tools used to curate the data, including the level of manual and automated practice, and how the processes, tools, and practices are documented. The applicant may find it useful for this particular Requirement to respond to each bullet point separately, and to address integrity and authenticity independently, as defined in the Guidance of the Requirement.

Audit trails which are written records of the actions performed on the data, should be described in the evidence provided.

VIII. 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 in determining whether data meet all criteria for inclusion in the collection and in establishing appropriate management for their preservation. Care must be taken to ensure that the data are relevant and understandable to the Designated Community served by the repository.

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?
- 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?)
- Does the repository have procedures in place to determine that the metadata required to interpret and use the data are provided?
- 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 quality control 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?

This Requirement addresses quality assurance from the viewpoint of the interaction between the depositor of the data and metadata and the repository. It contrasts with R11 (Data quality), which addresses metadata and data quality from the viewpoint of the Designated Community.

Extended Guidance R8.

The applicant should demonstrate that procedures are in place to ensure only data appropriate to the collection policy are accepted. Repository staff should have all the necessary information, procedures, and skills to ensure long-term preservation and use as applicable for the Designated Community.

IX. 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 with a preservation remit must also 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?
- What levels of security are required, and how are these supported?
- How is data storage addressed by the preservation policy?
- Does the repository have a strategy for backup/multiple copies? If so, what is it?
- Are data recovery provisions in place? What are they?
- 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?

This Requirement deals with high-level arrangements in respect of continuity. Please refer also to R15 (Technical infrastructure) and R16 (Security) for details on specific arrangements for backup, physical and logical security, failover, and business continuity.

Extended Guidance R9.

The reviewer will be looking to understand each of the storage locations that support curation processes, how data are appropriately managed in each environment, and that processes are in place to monitor and manage change to storage documentation. Can the repository recover from short-term disasters? Are procedures documented and standardized in such a way that different data managers, while performing the same tasks separately, will arrive at substantially the same outcome?

X. 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 legal 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:

- Is a preservation plan in place?
- Is the 'preservation level' for each item understood? How is this defined?
- 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?

Extended Guidance R10.

The reviewer will be looking for clear, managed documentation to ensure: (1) an organized approach to long-term preservation, (2) continued access for data types despite format changes, and (3) there is sufficient documentation to support usability by the Designated Community. The response should address whether the repository has defined preservation levels and, if so, how these are applied. The preservation plan should be managed to ensure that changes to data technology and user requirements are handled in a stable and timely manner.

If the applicant does not point to a preservation plan/policy, then they can be only at a maximum of compliance level 3 and should have one in place by the time of the next review.

XI. 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 work in concert with depositors to ensure that there is enough available information about the data such that the Designated Community can assess the substantive quality of the data. Such quality assessment becomes increasingly relevant when the Designated Community is multidisciplinary, where researchers may not have the personal experience to make an evaluation of quality from the data alone. Repositories must also be able to evaluate the technical quality of data deposits in terms of the completeness and quality of the materials provided, and the quality of the metadata.

Data, or associated metadata, may have quality issues relevant to their research value, but this does not preclude their use in science 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.
- Any automated assessment of metadata adherence to relevant schema.
- 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.

Provisions for data quality are also ensured by other Requirements. Specifically, please refer to R8 (Appraisal), R12 (Workflows), and R7 (Data integrity and authenticity).

Extended Guidance R11.

The applicants should make clear in the response that they understand the quality levels that can be reasonably expected from depositors. They should describe how quality will be assured during curation and the quality expectations of users, which may involve documentation of areas where quality thresholds have not been reached.

XII. 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 and reactive activities, archival workflows should be documented, and provisions for managed change should be in place. The procedure should be adapted to the repository mission and activities, and procedural documentation for archiving data should be clear.

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.
- Appraisal and selection of data.
- Approaches towards data that do not fall within the mission/collection profile.
- 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. Evidence of such workflows may have been provided as part of other task-specific Requirements, such as for ingest in R8 (Appraisal), storage procedures in R9 (Documented storage procedures), security arrangements in R16 (Security), and confidentiality in R4 (Confidentiality/Ethics).

Extended Guidance R12.

The reviewer is looking for evidence that the applicant takes a consistent, rigorous, documented approach to managing its activities throughout their processes and that changes to those processes are appropriately implemented, evaluated, recorded, and administered.

XIII. 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, and most repositories provide searchable catalogues describing their holdings such that potential users can evaluate data to see if they meet their needs. Once discovered, datasets should be referenceable through full citations to the data, including persistent identifiers to ensure that data can be accessed into the future. Citations also provide credit and attribution to individuals who contributed to the creation of the dataset.

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?
- 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?
- Does the repository offer persistent identifiers?

Extended Guidance R13.

The response should contain evidence that the curation of data and metadata is designed to support resource discovery of clearly defined and identified digital objects. It should be clear to users of the data how it must be cited to provide appropriate academic credit and linkages among related research.

XIV. 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 can be understood and used effectively into the future despite changes in technology. 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 required by the repository when the data are provided (e.g., Dublin Core or content-oriented metadata)?
- Are data provided in formats used by the Designated Community? Which formats?
- Are measures taken to account for the possible evolution of formats?
- Are plans related to future migrations in place?
- How does the repository ensure understandability of the data?

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.

Reuse is dependent on the applicable licenses covered in R2 (Licenses).

Extended Guidance R14.

The applicant should understand the needs of the Designated Community in terms of their research practises, technical environment, and applicable standards. Changes in technology are important, but appropriate high-quality metadata should also play an essential role and should be referred to in the evidence provided. The latter information is critical to design curation processes that result in digital objects meeting the needs of the end user, as well as generic or disciplinary standards.

Technology

XV. 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. Standards such as the OAIS reference model specify 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 (e.g., Spatial Data Infrastructure (SDI) standards, OGC, W3C, or ISO 19115)? 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.
- For real-time to near real-time data streams, is the provision of around-the-clock connectivity to public and private networks at a bandwidth that is sufficient to meet the global and/or regional responsibilities of the repository?

Extended Guidance R15.

The workflows and human actors providing repository services must be supported by a technological infrastructure. The reviewer is looking for evidence that the applicant understands the wider ecosystem of standards, tools, and technologies available for (research) data management and curation, and has selected options that align with local requirements. If possible, this should be demonstrated by using a reference model.

XVI. 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:

- Procedures and arrangements in place to provide swift recovery or backup of essential services in the event of an outage.
- Your IT security system, disaster plan, and business continuity plan; employees with roles related to security (e.g., security officers); and any risk analysis tools (e.g., DRAMBORA) you use.

This Requirement describes some of the aspects generally covered by others—for example, R12 (Workflows)—and is supplementary to R9 (Documented storage procedures).

Extended Guidance R16.

The reviewer is looking for evidence that the applicant understands the technical risks applicable to the servive for the data users and the physical environment, and that it has mechanisms in place to respond to security incidents. Evidence must focus on technical infrastructure rather than on managerial and procedural aspects of business continuity. In what way is the technical infrastructure controlled by the repository or by their host/outsource institution? Who is in charge? Can the repository in any way determine the technical infrastructure if that is outsourced? Are the arrangements sufficient to guarantee the long-term preservation of and/or access to the data holdings?

Applicant Feedback

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response

CoreTrustSeal is a community based non-profit organization promoting sustainable and trustworthy data infrastructures. This document consists of the Core Trustworthy Data Repositories Requirements for 2017–2019 and the Extended Guidance for CoreTrustSeal reviewers and applicants.

CoreTrustSeal
The Hague | +31 70 349 445
Tokyo | +81 4 232 763 95
E-mail: info@coretrustseal.org
Twitter: @CoreTrustSeal
Website: coretrustseal.org