Extended Guidance Version 2.0

CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022

~~Extended Guidance Version 1.1: June 2018~~

Contents

Introduction

This document ~~consists~~ contains the full text of the ~~**Core**~~ **CoreTrustSeal Trustworthy Data Repositories Requirements** for ~~2017–2019~~2020–2022 with introductory paragraphs on Background & General Guidance~~, which are set by~~.

In addition to the CoreTrustSeal ~~Board and~~ Requirements, which remain ~~unchanged~~stable for the period ~~2017–2019. The fixed text is recognizable by the boxes drawn around it.~~

~~The~~2020–2022, this document ~~furthermore contains~~provides the **Extended Guidance** for CoreTrustSeal reviewers and applicants. ~~This information will~~The Extended Guidance text may be ~~subject to change and updates~~updated during the period ~~for which the seal has been awarded, according to the needs of the community and~~2020–2022 subject to approval by the CoreTrustSeal Board. The document also contains a reference to the Glossary of Terms.

The document is intended to maximize consistency of reviews across the wide range of CoreTrustSeal applicants. The primary audience is reviewers, but it is also useful for applicants when preparing an application self-assessment.

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).

Response 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 Core Trustworthy Data Repositories Requirements Glossary: https://doi.org/10.5281/zenodo.3632563.

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~~Revising evidence information every three years for ~~a first round application should be supported by~~ only certification purposes is not efficient or effective. The ongoing ~~maintenance of publicly accessible business information. Revising all of your~~ management of business information ~~during the assessment is not what is being looked for. The (ongoing) management of information~~ necessary to run ~~your~~a repository's services should be sufficient to apply for and maintain certification. Specifically, repositories that document their policies and procedures well enough to ensure quality remains consistent, the risk of staff departure is mitigated, and so on, should need to only prepare application responses and manage public versions of their evidence.

~~Concerning the level of compliance, the reviewers~~ Reviewers may ~~decide the level of compliance is~~ propose a different ~~from the one~~ Compliance Level to that selected ~~for a Requirement~~by the applicant. 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~~a Requirement with a Compliance Level of less than 4 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~~It is not possible to cover every possible repository scenario in the ~~general guidance, each applicant is expected to extend their~~ Guidance or Extended Guidance. Likewise, not all bullet points in all Requirements are mandatory. Applicant responses should refer to the issues raised in the Guidance text and provide responses based on their local context. Final evaluation of a Requirement depends on the completeness and quality of the 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~~

Concepts and terminology used in the ~~self-assessment of a specific~~Requirements are informed by the OAIS Reference Model. The use of OAIS terminology can help to ensure understandability and clarity of the application, and applicants are therefore strongly encouraged to familiarize themselves with OAIS before preparing Requirement~~.~~ responses. The 2014 DPC Technology Watch Report 'The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)' by Brian Lavoie (https://doi.org/10.7207/twr14-02) provides a helpful introduction to the OAIS Reference Model.

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~~The CoreTrustSeal certification process~~, but all applicants will be judged against statements~~ depends on responses supported by ~~appropriate~~ evidence~~; not against self-assessed compliance levels'.~~

~~This means that specific assertions should be supported by links to~~ . The quality of public supporting evidence ~~as much as possible. In other words,~~is expected to increase over time. Applications are harder to assess if information is missing, insufficient, or unclear; if URL links are broken; or if 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~~statements continuously cross-reference one another ~~Requirement. Any deviations from this should be commented on explicitly.~~.

~~It also means that possible familiarity with~~Familiarity with, or inside knowledge of the repository by the reviewer ~~cannot~~must not play a role when judging the available evidence. The final, public evidence statement must also be clear for peer repositories to understand.

It is not expected for a reviewer to search through the applicant's website for evidence. If the information provided is insufficient for the reviewer to reach a decision, the application will be returned with an explanation of why the evidence is deficient. No Compliance Level will be given at this stage by the reviewer.

Understandability of documentation

Reviewers and readers of the final ~~review, which is made~~, public ~~on both the repository and the CoreTrustSeal websites,~~review should be able to understand the ~~procedure~~Requirement responses without detailed reading of supportive evidence. ~~For lengthy~~When longer documents are presented as evidence, or a document is used as evidence for more than one Requirement, the applicant should refer specifically to ~~the specific~~which sections ~~that provide the supporting evidence~~ are relevant 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~~their response.

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~~Sensitive and other internal documentation

CoreTrustSeal certification does not require supporting information to be made public 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 only on the intranet of ~~the~~a repository. Applicants may have ~~evidence documents containing such~~business information ~~alongside~~that contains both sensitive information *and* relevant evidence~~, and these should~~ for the CoreTrustSeal. Such evidence can be submitted confidentially to the reviewers and the documents named and described in the application[1]. ~~Applicants should aim at revising future documentation~~Over time we would expect applicants to ~~remove~~separate relevant evidence from confidential ~~information so that it can be made~~ materials, and assure a public ~~and considered during~~version is made available for the next review.

If documentation ~~is for internal use only, or if it~~ does not yet exist ~~or~~, is in progress, or is currently for internal use only (e.g., a wiki), then 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~~application. Certification may be ~~made in the comments by the reviewers, who will accept this~~approved on the ~~overall understanding that the documentation will be made public within the promised schedule. All deadlines will be part~~basis 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,~~these assurances. Applicants are expected to provide the public documentation when they ~~should be asked to publish at least a summary of the document~~renew their certification.

Application structure and length

~~The audience of CoreTrustSeal applications (i.e., self-assessments) is initially~~Applications in progress are confidential to 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~~. Even the Extended Guidance cannot cover every topic and evidence ~~required~~type that ~~is not covered by~~could be relevant to the ~~Guidance~~application. 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.

Applicants should not need to repeat long portions of text in different Requirement responses. In cases where evidence is applicable to more than one Requirement, a short summary statement of the relevant information should be added and then a cross-reference to the appropriate response for further details.

[1] ~~To safeguard the confidentiality of reviewers, confidential~~Confidential documents should ~~always~~ be submitted ~~via~~to the CoreTrustSeal ~~e-mail address.~~Secretariat via info@coretrustseal.org

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 (1) initial deposits are retained unchanged and that edits are only made on copies of those originals, and (2) metadata that enables the Designated Community to understand and use the data independently (i.e., without having to consult the original creator) is present at deposit or added by the repository. 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.

Extended Guidance R0.

Repository Type & Brief Description of Repository

Selection of more than one repository type should be supported by an explanation in *Brief Description of Repository* of how these multiple roles are fulfilled. This explanation could reference

relevant data collections, data types, formats, and disciplines the repository works with.

Brief Description of the Designated Community

As stated in the definition (see Glossary), it is possible for a repository to have a Designated Community composed of different 'sub-communities'; for example, for different collections. If this is the case for an applicant, they should provide a definition and sufficiently detailed description of each of these sub-communities. In addition, it is important to note that the Designated Community may be smaller than the overall group of users for a repository. The digital collections of a natural history museum may be appealing to a wide group of interested users, including the general public. Nevertheless, the museum may define its Designated Community as narrower than this (e.g., *biologists and anthropologists researching topics from the field of natural history*).

To serve its Designated Community well, a repository has to have a deep understanding of the Designated Community's composition, skills, knowledge base, and needs~~.~~, and how these may transform over time. Throughout the application, evidence should demonstrate an understanding of what the curation requirements are (additional context, preferred formats, etc.) to best serve the Designated Community (including respective sub-communities, if applicable), alongside a demonstration that the applicant monitors and responds to changes in the needs of the Designated Community.

A repository with a ~~very~~highly specific, narrow Designated Community might easily state the expected knowledge base (e.g., the degree~~/~~ of understanding of genetics, or the level of ~~understanding of statistics or genetics). A very~~expertise in using statistical software). In contrast, a broad ~~designated community (e.g., the general public) would imply that the repository has a wider~~Designated Community (i.e., composed of multiple user communities) means that the repository should have a sufficient understanding of all their knowledge bases and offer a wide range of contextual documentation to ensure its data can be understood by everyone~~.~~ within the Designated Community. With regard to defining the Designated Community's knowledge base, applicants should explicitly state any tacit assumptions, such as (foreign) language skills, ability to access specific Operating Systems or Internet browsers, use certain software, and so on.

Level of Curation Performed

More than one option (A, B, C, or D) of the level (or extent) of curation can be ~~picked~~selected, depending on the type of data and ~~agreement~~curation terms agreed with the depositor. When a repository performs curation at more than one level, further information should be added on the proportion of the data in the collection curated to the respective levels. Responses to the Requirements should then address how the curation workflows reflect the different curation levels, and how curation levels relate to preservation levels. For instance, are preservation goals and actions the same for all data, regardless of the curation level.

In addition to stating curation levels, a repository must therefore demonstrate that it assures long-term accessibility of data as the needs of the Designated Community change. This is less likely to be possible at curation levels A or B, because without normalizing submitted file formats to a common preservation format, it may be difficult to perform format migrations in the future depending on the heterogeneity of the collection. Similarly, lack of rich metadata and documentation may pose a risk concerning the continued usability of the data.

Reviewers will expect a higher level of formal provenance, integrity, and version management (change logs, etc.) ~~in the case of~~as curation levels progress from A through to D.

Insource/Outsource Partners

If more than one partner is involved, a diagram to indicate the full scale of the insourcing/outsourcing process is useful to assist the reviewers ~~in their understanding.~~. Having ~~more than one~~ multiple insourcing/outsource ~~partner~~partners (e.g., one for storage, one for maintaining websites) is ~~okay~~acceptable as long as all of the ~~relations~~relationships are clearly indicated~~, preferably in a diagram.~~. Reviewers will ask for the response in this section to be revised if the evidence statements later in the ~~self-assessment~~application 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. A clear and consistent description about the organizational approach as a whole is generally helpful.

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.

Extended Guidance R1.

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

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).

Extended guidance R2.

~~Access~~Stipulations on data access and use ~~conditions~~ could be defined in a set ~~differently: either as~~ of standard terms and conditions, or ~~as~~ differentiated ~~for~~by depositor or dataset. For sensitive data, in particular ~~depositors or datasets. These could cover the level of curation, what is the liability level, the level of responsibility taken for the data,~~, licenses may specify 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~~Popular license options include, but are not limited to, those offered by Creative Commons ~~licences~~ (~~https://creativecommons.org/), including~~ https://creativecommons.org/) such as 'CC 0 ~~Waiver~~Waiver' and 'public domain ~~data, could be used as a reference here, but other alternatives are also possible~~data' licenses.

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~~A Compliance Level of 4~~,~~ is necessary if the applicant is currently providing access to personal data.

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).

Extended Guidance R3.

The reviewer is looking for information to understand the level of responsibility taken for data~~.~~, the level of risk for the current organization, and the level of succession planning for the future of the data collection. For example, ~~are you~~is the applicant the primary or only custodian? ~~Is~~Does the depositor ~~responsible as well?~~share responsibility for the future of the data? 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.~~ temporary or permanent break in service.

~~IV~~The responsibility for sustainability may not lie in the hands of the repository itself, but with a higher host or parent organization. If so, this should be clearly indicated. Moreover, if the repository is part of a larger organization, has this or any other organization (e.g., a National Archive) guaranteed that it will take over the responsibility in the case of a service discontinuity? If there is no formal, written agreement between the repository and such an organization, then the Compliance Level can be at a maximum of 3 only.

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).

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, or protected areas and species, there are ~~in addition~~ additional legal and ethical 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).~~endangered species or an archaeological site). If there is any risk that identifiable data are deposited—for example, by accident—the repository must take appropriate measures for handling such data and to ensure they are dealt with (disposed of) in accordance with legal regulations.

~~Minimum compliance level should~~ The Compliance Level must be ~~a~~at 4 if the repository is currently providing access to personal or other sensitive data.

~~Reviewers expect to see evidence~~Evidence should demonstrate that the applicant understands their legal environment and the relevant ethical practices, and that they have documented procedures in place to ensure conformity.

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).

Extended Guidance R5.

The ~~description of~~response to 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 of funding occur?

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).

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~~community and technology watch). ~~Part~~If part of this information ~~may~~has already ~~have~~ been ~~given~~provided under 'R0. Brief Description of the Repository's Designated Community' and 'Other relevant information'~~. If so, then please refer to~~, the applicant should reference it.

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?

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,~~ ensure that data authenticity and integrity are protected throughout the entire curation lifecycle—including the level of manual and automated practice~~, —~~and how ~~the~~such processes, tools, and practices are documented. ~~The~~ As defined in the Guidance of the Requirement, 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.~~ (note that the response must be written in full prose).

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

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).

Extended Guidance R8.

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

For the collection to remain relevant to and usable by the Designated Community—particularly in light of changes in technology, culture, or legislation (e.g., data protection or intellectual property rights)—selection criteria may have to be revised over time and digital assets reappraised accordingly. Policies and documented procedures should be in place for the removal of items from a collection.

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).

Extended Guidance R9.

The reviewer ~~will be~~is 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? Examples of evidence include data flow diagrams covering deposit, curation, and access locations (plus any access restrictions). For archival storage, evidence might comprise of descriptions of the multisite arrangements (on site, near site, off site), the mix of storage media, and any redundancy (including integrity through checksums).

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).

Extended Guidance R10.

The term preservation plan refers to having a documented approach for defining and implementing preservation actions. The Requirements do not define or differentiate between a preservation policy, plan, strategy, or action plan.

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 preservation levels differ between classes or collections of items, the applicant should explain the differences in preservation approach, as well as the criteria applied to determine the preservation level. This may be relevant if, for example, the file size of an object or the sensitivity of the data it contains determines the number of redundant copies made; or, only items deposited in preferred formats are converted to standard preservation formats and will be migrated in the future.

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

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).

Extended Guidance R11.

The ~~applicants~~applicant should make clear in the response that they understand the quality levels that can be reasonably expected from depositors. ~~They~~Evidence should describe how quality will be assured during curation, and the quality expectations of ~~users,~~the Designated Community. Both the repository and its depositors are expected to document any areas in which ~~may involve documentation of areas where~~data or metadata quality ~~thresholds have not been reached.~~ falls below the expected standard.

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.

Extended Guidance R12.

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

The Requirement does not demand detailed descriptions of workflows, but seeks evidence of how and where these 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?

Extended Guidance R13.

The response should contain evidence that ~~the~~all curation of data and metadata ~~is designed to support resource~~ supports the discovery of digital objects that are clearly defined and identified, and enables their linkage with related digital objects in accordance with domain standards. It should be clear to ~~users of~~ the Designated Community how data ~~how it must be~~are cited ~~to provide~~such that appropriate ~~academic~~ credit and ~~linkages among related research.~~attribution is given to the individuals/organizations who contributed to their creation.

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.

Extended Guidance R14.

~~The~~ To meet this Requirement, the applicant should ~~understand~~demonstrate both an in-depth knowledge of reuse scenarios and the needs of the Designated Community in terms of their ~~research practises~~practices, technical environment, and (adherence to) applicable standards. Changes in technology and in the methodologies and norms employed by the Designated Community can lead to a need to reconsider the format in which data are ~~important, but~~disseminated. Similarly, appropriate, high-quality metadata ~~should also~~conforming to a generalized and/or disciplinary-specific schema 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~~ensure digital objects ~~meeting the needs of the end user,~~remain over time understandable and usable by the Designated Community. In particular, if only a generalized metadata schema (such as ~~well as generic or disciplinary standards.~~Dublin Core or DataCite) is employed, the applicant should provide evidence that this is sufficient for continued understandability of the preserved content by the Designated Community.

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).

Extended Guidance R15.

~~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 suitable technological infrastructure that meets the needs of the Designated Community and enables the repository to recover from short-term disasters. 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.

Examples of relevant standards include Spatial Data Infrastructure (SDI) standards, Open Geospatial Consortium (OGC), W3C, or ISO standards.

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?

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[2]) 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.

Extended Guidance R16.

The reviewer is looking for evidence that the applicant understands ~~the~~all technical risks applicable to the ~~servive for the data users and~~ service provided to the Designated Community, as well as to the physical environment~~, and~~. The applicant should demonstrate that ~~it has~~they have mechanisms in place to prevent, detect, and respond to a security ~~incidents. Evidence must focus on technical infrastructure rather than on managerial and procedural aspects of business continuity.~~ incident.

In what way is the security of 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?~~

Are authentication and authorization procedures in place sufficient to guarantee the security of the data holdings at each stage of the workflow (e.g., by requiring two-factor authentication for sensitive data)?

Which company security policies are in place to govern the security of all systems, including network security, intrusion checks, physical facility security, and password policy?

[2] 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