



**CoreTrustSeal Trustworthy Digital Repositories
Requirements 2023-2025**

Extended Guidance

V01.00

Table of Contents

Introduction	3
CoreTrustSeal Resources	3
Background & General Guidance	4
Compliance Levels	4
Supporting Evidence Links and Missing Information/Evidence	5
Internal Information, Sensitive Information & Confidentiality	5
Use of English, and non-English Language Documentation	6
Certification Validity & Renewal	6
Application structure and length	6
Requirements	7
R0. Background Information & Context	7
Organisational Infrastructure	11
Mission & Scope (R01)	11
Rights Management (R02)	12
Continuity of Service (R03)	13
Legal & Ethical (R04)	14
Governance & Resources (R05)	16
Expertise & Guidance (R06)	17
Digital Object Management	17
Provenance and authenticity (R07)	17
Deposit & Appraisal (R08)	18
Preservation plan (R09)	19
Quality Assurance (R10)	21
Workflows (R11)	22
Discovery and Identification (R12)	23
Reuse (R13)	23
Information Technology & Security	24
Storage & Integrity (R14)	24
Technical Infrastructure (R15)	25
Security (R16)	26
Applicant Feedback	27

Introduction

The CoreTrustSeal Standards and Certification Board¹, drawn from the CoreTrustSeal Community of Reviewers² manages the periodic revision of the CoreTrustSeal Trustworthy Repository Requirements, the peer review process and the final approval of certifications. It also seeks to contribute to and align with other organisations, standards and practices across the data management lifecycle.

An applicant for CoreTrustSeal must offer a long term preservation service. Some parts of the collection may have lower levels of care but this must be made clear in the text. Once assigned each reviewer will briefly check:

- the definition of the designated community to see whether it is clear enough.
- the preservation plan to ensure that active preservation is in place.
- the ingest & appraisal to confirm that digital objects receive active preservation.
- Reuse to confirm that the outcomes of curation are aligned with the needs of the designated community.

If it is not clear that the applicant offers active preservation, or it is not clear what other levels of curation are offered, or it is not clear that the designated community as defined is well served by the information in Reuse, then the applicant is either not in scope, or has provided insufficient information for a review to take place.

In this case the review will be returned to the applicant with comments on the relevant items for revision. These changes may imply other changes to the application at the applicants' discretion.

Applications under review 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.

Successful applicants can put staff members forward to become members of the community of reviewers. Members of this peer-review pool are eligible for Board membership.

In addition to the full CoreTrustSeal Requirements text, which remains stable for the period 2023-2025, this document provides the **Extended Guidance** for CoreTrustSeal reviewers and applicants. Extended Guidance text is presented within a border and in blue. This text may be updated during the 2023-2025 period.

CoreTrustSeal Resources

<https://www.coretrustseal.org/apply/>

<https://www.coretrustseal.org/why-certification/frequently-asked-questions/>

¹ <https://www.coretrustseal.org/about/standards-and-certification-board/>

² <https://www.coretrustseal.org/about/assembly-of-reviewers>

CoreTrustSeal Requirements v01.00 2023-2025

The full normative CoreTrustSeal Requirements and Guidance. Stable for the period 2023-2025.

CoreTrustSeal Extended Guidance v01.00 2023-2025

The full CoreTrustSeal Requirements text with extended guidance including comments and discussion. May be periodically updated during the period 2023-2025.

CoreTrustSeal Glossary v01.00 2023-2025

Definitions of key terms used in the CoreTrustSeal Requirements.

A number of the concepts and terms used in CoreTrustSeal are informed by the OAIS Reference Model³. Applicants are encouraged to familiarise themselves with this standard.

Background & General Guidance

The *CoreTrustSeal Requirements* describe the characteristics required to be a trustworthy repository for digital data and metadata. Each Requirement is accompanied by Guidance text describing the response statements and evidence that applicants must provide to enable an objective review. Applicants must respond to all of the Requirements.

Compliance Levels

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

- In Progress: the repository is in the implementation phase.
- Implemented: the requirement has been fully implemented by the repository.

Compliance levels are an indicator of the applicant's self-assessed progress, but reviewers judge compliance against response statements and supporting evidence.

A reviewer may reduce a compliance level to 'in progress' and provide an explanation to the applicant in feedback. All requirements assessed as 'in progress' must be supported by a statement from the applicant about the actions and timescales planned to reach 'implemented'. A reviewer will not increase a self-assessed 'in progress' compliance level to 'implemented'. Certification may be granted if some requirements are 'in progress'. When CoreTrustSeal is renewed, reviewers will expect to see a move from 'in progress' to 'implemented' or clear explanations as to why this is not possible.

During a renewal process an applicant may reduce a self-assessed compliance level from 'implemented' to 'in progress', e.g. if a significant upgrade is in place that has a temporary impact on the service. This level of transparency is highly desirable and, with sufficient explanation, should not be a barrier to renewing certification.

³Consultative Committee for Space Data Systems (June 2012). Reference Model for an Open Archival Information System (OAIS). Recommended Practice, issue 2, CCSDS 650.0-M-2. NASA.
<https://public.ccsds.org/pubs/650x0m2.pdf>

There is no formally applied maximum number of 'in progress' compliance levels that would stop an application from being successful. This will depend on the individual repository and the timescales and planning information provided for implementation. Reviewers will pay particular attention if a repository approach to continuity of service (R03) or active preservation planning (R09) is 'in progress'.

Supporting Evidence Links and Missing Information/Evidence

Response statements provided by applicants must be supported by links to public evidence online. Final versions of successful applications are public documents. This level of transparency is important, as the certification process does not include a site visit by an auditor. Links should be verified immediately before submitting applications.

Those reading applications (Reviewers, and eventually the public) should be able to understand the response statements without detailed reading of linked evidence. When longer documents are presented as evidence, or the same evidence is used to support more than one Requirement, the applicant must refer specifically to which sections are relevant and quote/summarise the information in their response.

The CoreTrustSeal certification process depends on applicant responses supported by clear evidence. The quality of public supporting evidence is expected to increase over time. Applications cannot be assessed if information is missing, insufficient, or unclear.

Prior knowledge of a repository by the reviewer must not play a role when assessing the applications. The final, public evidence statement must also be clear for peer repositories to understand.

A reviewer is not expected to search through the applicant's website for evidence. Applicants must provide specific references, including quotes/summaries of the cited information. The application will be returned with an explanation if the information provided is insufficient for the reviewer to reach a decision and assign a compliance level.

For evidence provided by a party other than the applicant, the relationship with that party should be described, see *Cooperation and outsourcing to third parties, partners and host organisations*.

Internal Information, Sensitive Information & Confidentiality

No sensitive information disclosure is required to acquire CoreTrustSeal. If evidence cannot be made public it is possible to share this confidentially during the certification process.

CoreTrustSeal certification does not require supporting evidence to be made public that is confidential, commercially sensitive, or poses a security risk. Applicants may have internal business information that contains both sensitive information *and* relevant evidence for the

CoreTrustSeal. Such evidence can be submitted confidentially to the reviewers⁴ and the documents named and described in the application. Over time, applicants should separate relevant evidence from confidential materials, and assure a public version is made available for the next review.

If documentation does not yet exist, is in progress, or is currently for internal use only, then a date of public availability should be stated in the application. Certification may be approved based on these assurances. Applicants are expected to provide the public documentation when they renew their certification.

Use of English, and non-English Language Documentation

All responses must be in English. If links to non-English evidence are provided, then an English summary must be included in the response statement. This summary can be brief for certain types of documents (e.g. a reference to a list of preferred formats), but should be longer for others (e.g. a Preservation Policy document).

Full English translations of linked evidence are not required.

Certification Validity & Renewal

CoreTrustSeal certification is valid for three years from the date of certification. An organisation with well-managed business processes and records should be able to reapply with minimal revisions. More significant revisions may be required if:

- the organisation, its data collection, technical infrastructure or Designated Community changes significantly
- the CoreTrustSeal Requirements are 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.

Application structure and length

It is not possible to cover every possible repository scenario in the Guidance or Extended Guidance and some guidance or questions may not be locally applicable. 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. Reviewers are looking for clear, open statements of evidence specific to the applicant. It is understood that the length of response statements will vary, but the overall application should provide a focussed narrative describing the supporting evidence.

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.

⁴ Contact the CoreTrustSeal Secretariat via info@coretrustseal.org

Reviewers understand that applicants' organizational structures, missions, size and digital object collections vary widely. Even the Extended Guidance cannot cover every topic and evidence type that could be relevant to the application. Some additional text may be needed to explain the relevance of evidence provided; especially, if not available in English. No minimum or maximum lengths for responses are defined, but even the most complex evidence statements are usually at the lower end of the 500–800 word range. Evidence statements should be supported by public links to the documentation the applicant uses to manage their organization and digital objects. It is this public evidence that offers the most assurance of compliance with the Requirements.

The CoreTrustSeal Requirements seek to minimise repetition and overlap, but some applicants may submit the same evidence for more than one requirement. 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 can be provided with a reference to the Requirement that contains further details.

Requirements

R0. Background Information & Context

This section provides the information necessary for reviewers to fully assess the applicants response statements. It is important to the entire application that the correct options are selected and that sufficiently detailed responses are provided.

(1) *Re3data Identifier*⁵. |

Response

(2) *Repository type*. | Select a repository type:

- **Generalist repository**
- **Specialist repository**
 - Specialist repositories are asked to provide their domain(s) and/or discipline(s).

Response

As stated in the glossary (ref), a specialist repository is a domain or subject-based repository which specializes in a specific (research) field or data type, and supports that defined designated community. A generalist repository does not specialise in a domain, discipline, specific (research) field or data type and supports a defined designated community.

(3) *Overview*. Provide a short overview of key characteristics of the repository, reflecting the

⁵ <https://www.re3data.org/>

repository type selected. This should include information about the scope and size of data collections, data types and formats. Further contextual information may also be added.

The overview should include contextual information that is not covered elsewhere in the Requirements.

Response

(4) 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 group(s) of users at whom the curation and preservation measures are primarily targeted. The definition should be specific so that reviewers can assess whether that community is being served in the responses to other requirements.

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, the applicant should provide a definition and sufficiently detailed description of each of these sub-communities. It is important to note that the Designated Community may be smaller than the overall group of consumers of the repository data, metadata and services. 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*).

A repository must have an understanding of the Designated Community’s composition, skills, knowledge base, and needs, and how these may transform over time. This includes an understanding of typical re-use scenarios and purposes, whether they are as general as “Read online publications on a computer to learn more about the history of X” or as specific as “Run statistical analyses using SPSS”. Throughout the application, evidence should demonstrate an understanding of what the curation and preservation actions (additional context, preferred formats, etc.) will best serve the Designated Community (including respective sub-communities, if applicable). It should also be clear how the applicant monitors and responds to changes in the needs of the Designated Community.

A repository with a highly specific, narrow Designated Community might easily state the expected knowledge base (e.g., the degree of understanding of genetics, or the level of expertise in using statistical software). In contrast, a broad 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.

Response

(5) Levels of Curation.

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 during ingest, enhancement of documentation and metadata**
- D. Data-level curation – as in C above, but with additional editing of deposited data**

Response

Guidance

A repository must demonstrate that it assures long-term accessibility and understandability 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 normalising 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.

It is recognised that a repository may offer different levels of curation to different digital objects. It is important that this is clear to depositors, users, and to CoreTrustSeal Reviewers.

More than one option (A, B, C, or D) of the level (or extent) of curation can be selected, depending on the type of data and curation terms agreed with the depositor. For each level selected add some concise information on how the respective levels are reached e.g. automatic checks of metadata, intellectual checks and editing of documentation, file format identification, transformation to preservation file formats, etc.

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. In this case, applicants should take care that responses to the Requirements state any relevant differences in workflows or employed measures for each selected curation level.

All levels of curation assume (1) initial deposits are retained unchanged and that edits are only made on copies of those originals, (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, and (3) ongoing measures for active preservation are in place for the greater part of the collection(s).

Annotations/edits must fall within the terms of the license agreed with the data depositor 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.

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

(6) Cooperation and outsourcing to third parties, partners and host organisations.

Response

Guidance

If the applicant is entirely responsible for all decisions and takes all relevant actions related to meeting each of the 16 Requirements then this section can be left blank. If for one or more requirements the applicant is supported by another organization in making decisions or taking actions, that organisation, the role it plays, and its relationship with the applicant should be listed here.

It is understood that repositories may be structured in different ways. It is important that repository certification is associated with a clearly defined organisation. The structure of the applicant organisation is addressed under Governance & Resources (R05).

If a repository function and/or supporting evidence that is covered by the CoreTrustSeal is not under the direct control of the applicant, then the relevant host organisation, partner or other third party should be listed here. Describe the function or service they provide, the nature of the relationship or agreement (contractual, Service Level Agreement, Memorandum of Understanding, etc.) and whether they have any relevant certifications. Appropriate qualifications or certifications, including but not limited to the CoreTrustSeal, are preferred but not required. Explaining what types of agreement are in place, or why these may not be practical, helps ensure transparency. It is not expected that applicants share commercial or otherwise sensitive details of relationships (see: Internal Information, Sensitive Information & Confidentiality).

Such relationships may include, but are not limited to: cooperation or federation with other repositories, any services provided by an institution the applicant is part of, storage provided by others as part of multi-copy redundancy, or organisations that may undertake some responsibility for data, metadata and services in a service continuity or succession situation.

The listed organisations should then be clearly referenced under each relevant Requirement.

Because outsourced functions will usually still have some level of shared responsibility the applicant must provide appropriate evidence for Requirements that are not outsourced, and for the parts of the data lifecycle that they control.

Though a wide range of services and functions may be outsourced, a CoreTrustSeal applicant must retain responsibility for the preservation planning and actions

undertaken to data and metadata to ensure they remain usable by their Designated Community for the long term.

This can be a complex area to define and describe, but such details are essential to ensure a comprehensive review process.

If more than one partner is involved, a context diagram (e.g. the OAIS functional model) to indicate the full scale of the outsourcing process is useful. Having multiple partners (e.g., one for storage, one for maintaining access systems) is acceptable as long as all of the relationships are clearly indicated. Reviewers will ask for the response in this section to be revised if the evidence statements later in the application refer to third parties, partners or host organisations not mentioned here.

(7) Applicants renewing their CoreTrustSeal certification: summary of significant changes since last application. CoreTrustSeal certification has an expectation of continuous improvement over time. Repositories undergoing recertification should highlight briefly any significant changes including to technical systems, Designated Community or funding during the previous three years. This could include any steps taken to move from 'In Progress' to 'Implemented' Requirements since the last certification.

Response

Organisational Infrastructure

Mission & Scope (R01)

R01. The repository has an explicit mission to provide access to and preserve digital objects.

Self-Assessed Compliance Level:

Response

Guidance

Repositories take responsibility for the curation of digital objects, and for ensuring that materials are held in the appropriate environment for appropriate periods of time. For Trustworthy Repositories it must be clear to depositors and users that active preservation of and continued access to the digital objects is an explicit role of the repository.

The response statement and evidence should include references to the following items:

- The mission to actively preserve and provide access to digital objects
- The level of approval that the mission has received.

Evidence for this Requirement could include an approved public mission statement, roles mandated by funders, or a policy statement signed off by a governing board.

If preservation is not referred to in the mission of the repository or other relevant public documents provided as evidence, then the compliance level cannot be higher than “In Progress: the repository is in the implementation phase”.

Rights Management (R02)

R02. The repository maintains all applicable rights and monitors compliance.

Self-Assessed Compliance Level:

Response

Guidance

The repository manages, and communicates to relevant stakeholders, all rights (permissions, prohibitions, obligations) covering data and metadata deposit, storage, preservation, access, and use.

This requirement relates to the system, methods and artefacts (e.g. licenses, agreements, terms and conditions, and related policies and procedures) in place for rights management.

The repository must obtain all necessary rights from the depositor, and demonstrate that there are sufficient controls in place to ensure they are applied and monitored.

The response statement and evidence should include references to the following items:

- The overall rights management approach to deposited files, data and metadata.
- The rights to copy, transform, and store digital objects for preservation, as well as provide access to them
- Conditions of use (e.g. intellectual property rights, distribution, intended use, protection of sensitive data, etc.).
- Deposit and access agreements or licenses.
- How rights metadata is managed for humans (e.g. license documents/files) or machines.
- Monitoring of compliance at deposit, during curation/preservation, and during access and reuse. Describe any circumstances where compliance monitoring is not possible.
- Measures in place if non-compliance is detected.

Data and metadata, including ‘open data’, will usually have some rights attached even if there is no signed license artefact or formal agreement in place. This could include obligations such as citation and attribution of data and metadata used, or making secondary analysis openly available. If all data and metadata are made available without any conditions of access or use then this should be made clear in the response statement.

Rights negotiations and transfer should be described under Deposit & Appraisal (R08). Any ethical codes of conduct, privacy measures, or legislation that influence rights management should be described under Legal & Ethical (R04).

Stipulations on data access and use could be defined in a set of standard terms and

conditions, or differentiated by depositor or type of digital object. Popular licence options include, but are not limited to, those offered by Creative Commons (<https://creativecommons.org/>) such as 'CC 0 Waiver' and 'public domain data' licences.

For sensitive data, in particular, licences may specify limitations on use, usage environment (safe room, secure remote access), and types of users (approved researcher, minimum training requirements, etc.). More recently, the Local Contexts Traditional Knowledge and BioCultural Notices and Labels have emerged as a means for Indigenous peoples to allocate customizable statements of provenance, protocols and permissions to research objects.⁶

While it may be challenging to identify instances of noncompliance, consideration should be given to the consequences if noncompliance is detected e.g. sanctions on current or future access/use of data and metadata. 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.

For applicants that hold data or metadata with a disclosure risk, the target compliance level should be "Implemented: the requirement has been fully implemented by the repository".

Continuity of Service (R03)

R03. The Repository has a plan to ensure ongoing access to and preservation of its data and metadata.

Self-Assessed Compliance Level:

Response

Guidance

The repository must have measures in place to address the risks inherent in changing circumstances, including in mission and/or scope. This Requirement covers the stable management of repository services over time (business continuity) and the response when services have problems (disaster recovery). It also includes preparations for handover of digital objects and services to another repository (succession planning). The deposit, storage, preservation, and access services offered by the repository to depositors and users are all in scope.

The response statement and evidence should include references to the following items:

- The functions and services offered by the repository to depositors and users.
- The approach to rapid changes of circumstance and long-term planning.
- The options for relocation or transition of the activity to another repository. For example, the case of cessation of funding due to an unexpected withdrawal of funding, or a shift of host institution interests.

⁶ <https://localcontexts.org/>

- The repository approach to managing policies, procedures and other business information over time.

Even though succession agreements may be hard to achieve it is important to acknowledge the possibility that a repository will cease to function or exist. If there is no formal, written agreement between the repository and a successor then the compliance level cannot be higher than “In Progress: the repository is in the implementation phase”.

Any technical aspects of business continuity, and disaster and succession planning should be covered in R15 (Technical infrastructure).

Repositories must ensure continuity of their collections and assume responsibility in the case of a temporary or permanent break in service. Responses and evidence should demonstrate the level of responsibility taken for digital objects, the level of risk for the repository, and the level of succession planning for the future of the data collection.

Relevant information could include whether the applicant is the primary or only custodian, whether the depositor shares some responsibility for the future of the digital objects and any service level guarantees or minimum guaranteed time periods (e.g. for retention or preservation) in place.

If sustainability partially depends on a host or parent organisation, or another organisation has guaranteed that it will take over the responsibility in the case of a service discontinuity, this should be clearly indicated. Identifying and entering a formal agreement with a successor organisation that can undertake to deliver the same levels of care and service is acknowledged as a challenge for many repositories. For this reason a continued status of ‘in progress’ may be accepted as sufficient during renewal of certification if clearly explained.

If there is no formal, written agreement between the repository and a successor organisation this requirement cannot be assessed as ‘implemented’.

Legal & Ethical (R04)

R04. The repository ensures to the extent possible that data and metadata are created, curated, preserved, accessed and used in compliance with legal and ethical norms.

Self-Assessed Compliance Level:

Response

Guidance

This requirement relates to repository awareness and processes around legal and ethical issues, including privacy and confidentiality, that impact the creation, curation, and use of digital objects.

To maintain the trust of those who agree to have their digital objects held by the repository,

evidence should demonstrate practices that reflect the legal status and sensitivity of digital objects, including guidance for depositors and users.

The response statement and evidence should include references to the following items:

- How the repository identifies and manages relevant legal and ethical standards that impact operations.
- Compliance with specific legal and/or ethical discipline or domain standards.
- Information requested from depositors to confirm that data collection or creation was carried out in accordance with legal and ethical criteria in the relevant geographical location or discipline (e.g. Ethical Review Committee/Institutional Review Board or Data Protection legislation).
- Any data or metadata with disclosure risk e.g. depositor/user information, personal, cultural, or environmental information

For applicants that hold data or metadata with disclosure risk include references to the following items:

- Special procedures applied to manage disclosure risk
- Conditions of distribution, access protection and use
- Processes to review disclosure risk and to take the necessary steps to either anonymize files or to provide access in a secure way
- Staff training in the management of digital objects with disclosure risk.
- Guidance provided on the responsible deposit, download, and use of disclosive or potentially disclosive data and metadata.

The management of related rights and compliance checks should be covered under Rights (R02). Measures to protect digital objects should be addressed under Security (R16).

All organizations responsible for data have an ethical duty to manage them to the level expected by the Designated Community. In addition to national and international expectations of scientific practice, repositories holding data about individuals, organizations, indigenous peoples or protected areas and species, there are additional legal and ethical expectations.

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 identification of a person who participated in a survey or the location of endangered species or an archaeological site. If there is any risk that identifiable data are deposited the repository must take appropriate measures to ensure they are dealt with in accordance with legal regulations.

For applicants that hold data or metadata with a disclosure risk, the target compliance level should be "Implemented: the requirement has been fully implemented by the repository".

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. This requirement includes the appropriate handling of data and metadata about the users of repository services.

Governance & Resources (R05)

R05. The repository has adequate funding and sufficient numbers of staff managed through a clear system of governance to effectively carry out the mission.

Self-Assessed Compliance Level:

Response

Guidance

This Requirement reflects a need for transparency of financing, governance, responsibilities, and decision making. Evidence should demonstrate that the repository has a clear system of governance and sufficient human and financial resources to carry out its mission.

The response statement and evidence should include references to the following items:

- Descriptions and diagrams of governance bodies, groups and hierarchies.
- Timescales for provision and renewal of funding for operational costs and recruitment; it is understood that permanent, ongoing funding cannot be perfectly quantified or guaranteed.
- Evidence that the repository is, or is hosted by, a recognized institution (supporting long-term stability and sustainability) appropriate to its Designated Community.
- Demonstrate that the repository can meet its obligations, including sufficient funding, staff resources, IT resources, and a budget for external engagement when necessary.

The availability of appropriate expertise is covered under Expertise (R06) below.

Responses and evidence should demonstrate that the repository can meet its obligations, including sufficient funding, staff resources, IT resources, and a budget for external engagement when necessary. The organization's governance/management decision-making processes and the entities involved should be clear e.g. through organizational diagrams.

Evidence should include how often periodical renewal of funding occurs. It is acknowledged that repositories work under a range of different funding models, and often with limited resources, but demonstrating awareness of these issues is important to trustworthiness. It is acknowledged that past funding is not a guarantee of future funding, but it may be indicative to state the historical timescale of the repository.

Other relevant information could include the balance of structural versus project funding, the total number of full time equivalent (FTE) employees, and the proportions of staff employed on a permanent or temporary basis.

Expertise & Guidance (R06)

R06. The repository adopts mechanisms to secure ongoing expertise, guidance and

feedback-either in-house, or external.

Self-Assessed Compliance Level:

Response

Guidance

A repository must identify the skills necessary to deliver the services it offers, and source and maintain those skills either as internal resources or through external engagement. 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.

The response statement and evidence should include references to the following items:

- That guidance and expertise reflects the scientific scope of the repository, if relevant.
- The repository aligns internal recruitment and external engagement with the services it offers.
- The repository ensures that its staff have access to ongoing training and professional development.
- The range and depth of expertise of both the organisation and its staff, including any relevant affiliations (e.g. national or international bodies), is appropriate to the mission.
- In-house advisers, or external advisory committees that include technical, curation, data science, data security, and disciplinary experts.
- How the repository communicates with experts for advice.

Responses and evidence should demonstrate that the repository has sufficient internal expertise and is linked to a wide network for advice and guidance. Evidence must account for the repository day-to-day activities and the monitoring of potential new challenges on the horizon (community and technology watch).

Digital Object Management

Provenance and authenticity (R07)

R07. The repository guarantees the authenticity of the digital objects and provides provenance information.

Self-Assessed Compliance Level:

Response

Guidance

The repository should provide evidence to show that it operates a data and metadata management system that maintains provenance information to ensure authenticity from deposit, and through curation and preservation to the point of access.

Any intentional changes to data and metadata should be documented, including the rationale and originator of the change. Authenticity covers reliability and provenance, including the relationship between the deposited digital objects and those provided at the point of access.

The response statement and evidence should include references to the following items:

- The repository approach to changing and versioning data and metadata. How the approach and records of changes are communicated to data depositors and users.
- The provenance information and audit trails recorded for data and metadata processing and versioning.
- How the repository compares the essential properties of different versions of the same file.
- Identification checks for depositors.

Responses and evidence should provide a clear overview of the processes used to ensure data authenticity throughout the entire curation and preservation lifecycle—including the level of manual and automated practice.

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

Deposit & Appraisal (R08)

R08. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for users.

Self-Assessed Compliance Level:

Response

Guidance

The appraisal function during deposit is critical to evaluate whether digital objects meet all criteria for selection and to ensure appropriate management for their preservation. Appraisal ensures that deposited digital objects are relevant and are, or can become, understandable to the Designated Community.

The response statement and evidence should include references to the following items:

- Any documented deposit process that includes steps to ensure that data and metadata are sufficient for long-term preservation.
- A collection development policy or procedures to guide the selection of digital objects.
- Criteria for prioritisation and any different curation-levels or preservation levels defined during appraisal.
- The approach to digital objects that do not fall within the mission/collection profile.
- Procedures to determine that the metadata required to interpret and use the digital

objects are provided.

- Any automated assessment of metadata adherence to relevant schemas.
- The repository approach if metadata provided is insufficient for long-term preservation.
- A list of preferred formats.
- Checks in place to ensure that depositors adhere to the preferred formats.
- The approach towards digital objects that are deposited in non-preferred formats.
- The transfer of custody and responsibility during the handover from the depositor to the repository.

This Requirement covers the selection criteria applied at the point of deposit. Data Quality (R11) should be used to address steps taken by the repository during the curation process.

Responses and evidence should demonstrate that only data and metadata appropriate to the documented collection development policy or procedures are accepted. Repository staff should have all the necessary information, procedures, and expert knowledge to ensure long-term preservation and use as applicable to 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 ; this may influence reappraisal (see Preservation Plan- R09).

Preservation plan (R09)

R09. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Self-Assessed Compliance Level:

Response

Guidance

The repository, depositors, and Designated Community need to understand the level of responsibility undertaken for the long-term preservation of data and metadata. Procedures must be documented and their completion assured.

The response statement and evidence should include references to the following items:

- The documented approach to preservation, including whether this involves format migration, emulation, etc. Ensuring bit level integrity is vital but not sufficient for preservation.
- File formats and metadata schemas for long term preservation.
- How the level of responsibility for the preservation of each item is defined.
- Plans related to future migrations or similar measures to address the threat of obsolescence.
- Actions relevant to preservation specified in documentation, including custody transfer,

submission information criteria, and preservation information metadata.

- Measures to ensure these actions are taken.
- Any minimum stated retention and/or preservation periods.
- How often the digital objects are re-appraised and the possible outcomes of reappraisal.
- The repository approach to deleting/removing data and metadata from collection/holdings including the impact on persistent identifiers.

The rights of the repository, including the right to preserve, are covered under Rights Management (R02). Bit level integrity is covered under Storage and Integrity (R14). Acceptable file formats at deposit should be covered under Deposit and Appraisal (R08). Measures to ensure that file formats, schemas and content are appropriate to the Designated Community should be covered under Reuse (R13).

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.

Responses and evidence should demonstrate 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, metadata, technology and user requirements are handled in a stable and timely manner.

If preservation levels differ between classes or collections of items, the differences in preservation approach, and the criteria applied to determine the preservation level should be explained. 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.

Policies and documented procedures for reappraisal should be in place to manage changes to the curation or preservation levels of digital objects, or their deletion or removal from the repository.

Applications that do not link to a documented preservation approach can be only at a maximum Compliance Level of 'In Progress'. There must be a link to a documented plan by the time of renewal.

Quality Assurance (R10)

R10. The repository addresses technical quality and standards compliance, and ensures that sufficient information is available for end users to make quality-related evaluations.

Self-Assessed Compliance Level:

Response

Guidance

Different repositories undertake different levels of curation on data, metadata and documentation depending on the needs and expectations of their depositors and Designated Community. Quality assurance by the repository ensures that digital objects comply with a range of standard criteria including acceptable formats, metadata schema, metadata content and links to other digital objects. This relates to 'technical quality' rather than the 'scientific quality' of the original digital objects creation or collection prior to deposit, though the repository must ensure there is sufficient information about the digital objects for the Designated Community to assess their fitness for use. 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.

The response statement and evidence should include references to the following items:

- The approach to data and metadata quality taken by the repository including variations for different curation-levels.
- The standards that data, metadata and documentation must comply with to be acceptable for preservation and access. Whether these are general external standards, internally developed standards or specific to a community of practice.
- The quality control checks in place ensure the completeness and understandability of data and metadata.
- The approach to resolving issues e.g. whether the digital objects are returned to the depositor for rectification, fixed by the repository, noted by quality flags, and/or included in the accompanying metadata.
- The approach to managing changes to expected standards (e.g. new or updated data formats of metadata schemas) in response to changes in the technical environment or to changes in the needs of the Designated Community.
- Any links provided to other digital objects' data and metadata e.g. related digital objects, publications, or the use of controlled vocabularies and ontologies.

This Requirement refers to data and metadata quality standards and assurance during curation. Selection criteria are covered during Deposit and Appraisal (R08). Measures to ensure that digital objects remain fit for purpose over time are covered under Preservation Plan (R09).

Responses and evidence should demonstrate an understanding of the quality levels that can be reasonably expected from depositors. Evidence should describe how quality will be assured during curation, and the quality expectations of the Designated Community. Both the repository and its depositors are expected to document any areas in which data or metadata quality falls below the expected standard.

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.

Workflows (R11)

R11. Digital object management takes place according to defined workflows from deposit to access.

Self-Assessed Compliance Level:

Response

Guidance

For Quality Assurance (R10) to be achieved, it is necessary to avoid ad hoc actions and to deliver consistency of practice for all digital objects and across repository functions. This requires that workflows be defined, documented, and change-managed. Workflows may be specified in a mixture of standard operating procedures, business process descriptions and diagrams that guide normal practice and provide mechanisms for handling exceptions.

The response statement and evidence should include references to the following items:

- Workflows/business process descriptions covering the curation levels performed.
- How workflows are adjusted for different types of data and metadata.
- Decision handling within the workflows.
- Change management of workflows.
- Ability to track workflow execution, with mechanisms to handle exceptions.

This Requirement confirms that all workflows are documented. It should be noted if there are different workflows for different levels of security mentioned in the Legal and Ethical (R04) response statement. Workflows may include qualitative and quantitative checking of outputs, but any detail on checks and compliance should be addressed under Quality Assurance (R10).

Responses and evidence should demonstrate a consistent, rigorous, documented approach to managing all activities throughout their processes and that changes to those processes are appropriately implemented, evaluated, recorded, and administered.

Detailed descriptive workflows may be linked to as evidence if they are public, but this Requirement is seeking assurance that these workflows exist and that they are well documented. Process diagrams or decision trees may be used to illustrate workflows. Highlighting the relationships between processes from the other requirements can help to explain how individual workflows build into an overall system. .

Discovery and Identification (R12)

R12. The repository enables users to discover the digital objects and refer to them in a persistent way through proper citation.

Self-Assessed Compliance Level:

Response

Guidance

Effective data and metadata sharing discovery is key to resource discovery. Once discovered, digital objects should be referenceable through full citations, including persistent identifiers (PIDs) to help ensure that they can be accessed into the future.

The response statement and evidence should include references to the following items:

- The search facilities offered by the repository.
- The standards that a searchable metadata catalogue complies with.
- The approach to ensuring that identifiers are unique and persistent.
- Machine harvesting of the metadata.
- Repository, or repository data and metadata, inclusion in disciplinary or generic registries of resources.
- Recommended data citations.

Applicants should describe their use of a third party persistent identifier system, or document their own approach to ensuring that identifiers remain globally unique and persistent. The use of a third party to support PID creation and resolution is not sufficient; applicants should describe how they ensure that identifiers continue to resolve to the correct data or metadata over time, including the version rules that guide when a new identifier is created for a digital object. Applicants that do not have a persistent identifier solution cannot achieve “Implemented: the requirement has been fully implemented by the repository” for this requirement.

Responses and evidence should demonstrate that all curation of data and metadata supports the discovery of digital objects that are clearly defined and identified, and enables their linkage with related digital objects in accordance with relevant standards (e.g. discipline or domain). Attribution and citation guidance should be provided to users to ensure that appropriate credit is given to the individuals/organizations who contributed to the collection and creation of data and metadata.

Reuse (R13)

R13. The repository enables reuse of the digital objects over time, ensuring that appropriate information is available to support understanding and use.

Self-Assessed Compliance Level:

Response

Guidance

Repositories must ensure that data and metadata continue 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 and metadata are reusable.

The response statement and evidence should include references to the following items:

- The ways in which the repository engages with their Designated Community of users to identify their needs.
- The data formats, metadata schemas, controlled vocabularies and ontologies used to support reuse, and how these meet the community needs.
- The metadata and documentation provided at the point of access to support understandability and reuse appropriate to the Designated Community. This may include information specific to data type, e.g. manuals, calibration records, photos, protocols.
- Measures to ensure that data and metadata remain understandable.
- Management of changes to data, metadata, documentation or other information that supports reuse.

Responses to this Requirement should focus on engagement with the Designated Community, identification of their needs and specifying how their needs are met.

Responses and evidence should demonstrate both an in-depth knowledge of reuse scenarios and the needs of the Designated Community in terms of their 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 change the structure, content or delivery mechanisms for data and metadata. Appropriate, high-quality metadata conforming to a general and/or disciplinary-specific schema should be referred to in the evidence provided. This information is critical to the design curation processes that ensure digital objects remain understandable over time and usable by the Designated Community. If only a general level metadata schema (such as Dublin Core or DataCite) is in place, the evidence should demonstrate that this is sufficient for continued understandability of the preserved content by the Designated Community.

Information Technology & Security

Storage & Integrity (R14)

R14. The repository applies documented processes to ensure data and metadata storage and integrity.

Self-Assessed Compliance Level:

Response

Guidance

In addition to maintaining 'archival' copies of digital objects, repositories need to store data

and metadata from the point of deposit, for curation and preservation, and for access by users. For each storage location, measures should be in place to ensure that unintentional or unauthorised changes can be detected and correct versions of data and metadata recovered.

The response statement and evidence should include references to the following items:

- Processes and documents to ensure that the repository staff have a clear understanding of all storage locations and how they are managed.
- The repository's strategy for multiple copies.
- The risk management techniques used to inform the strategy.
- Procedures for handling and monitoring deterioration of storage media.
- Procedures to ensure that data and metadata are only deleted as part of an approved and documented process.
- Any checks (i.e. fixity checks) used to verify that a digital object has not been altered or corrupted from deposit to use.

Storage and integrity measures should be covered here (R14) and not as part of Technical Infrastructure (R15) or Security (R16) responses. Details of how intentional changes to the data and metadata are logged should be covered under Provenance & Authenticity (R07).

Responses and evidence should cover each of the storage locations that support curation processes, how data and metadata are managed in each environment, and that processes are in place to monitor and manage changes to storage documentation. Standard operating procedures should be sufficient that different storage managers will arrive at substantially the same outcome, even if performing the same tasks separately. Examples of evidence include data flow diagrams covering deposit, curation, and access locations (plus any access restrictions). For archival storage, evidence might include descriptions of the multisite arrangements (on site, near site, off site), the mix of storage media, and any redundancy (including integrity through checksums).

Technical Infrastructure (R15)

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

Self-Assessed Compliance Level:

Response

Guidance

Repositories must operate on reliable and stable core infrastructure that maximises service availability. The details of technical infrastructure will vary widely across repositories. Responses and evidence should focus on demonstrating that the repository solution, including hardware and software is well managed and appropriate to the needs of the repository functions and the Designated Community of users.

The response statement and evidence should include references to the following items:

- The repository software used for deposit, curation, preservation and access management. Whether it is community supported, open source, or locally developed.
- Any IT service management approach followed and the functions this approach specifies (e.g. systems documentation, software inventories, code repositories, infrastructure development planning).
- Any international, community or other technical infrastructure standards in place and how compliance is monitored.
- The version control systems used for repository generated software.
- Measures taken to ensure that availability, bandwidth, and connectivity are sufficient to meet the needs of the Designated Community.
- Processes in place to monitor and manage the need for technical change, including in response to the changing needs of Preservation (R10), and Reuse (R13) by the Designated Community.

Technical aspects of business continuity, disaster recovery and succession planning are relevant here, but their management should be covered under Continuity of Service (R03). This requirement excludes Security (R16) measures and Storage & Integrity (R14).

File formats and metadata schema information should be referenced under Deposit & Appraisal (R08) and Reuse (R13). Standards that are not technical or security focussed should be referenced under Quality Assurance (R10).

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. Responses and evidence should demonstrate an understanding of the wider ecosystem of standards, tools, and technologies available for (research) data management and preservation.

It should be made clear that the selected technical options align with local requirements e.g. technologies used by the designated community, or bandwidth that is sufficient to meet the predicted demand.

Examples of relevant technical standards include W3C, ISO, and IEEE standards.

Security (R16)

R16. The repository protects the facility and its data, metadata, products, services, and users.

Self-Assessed Compliance Level:

Response

Guidance

The repository should analyze potential threats, assess risks, and create a consistent security system. It should consider damage scenarios based on malicious actions, human error, or

technical failure that pose a threat to the repository and its data, metadata, 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.

The response statement and evidence should include references to the following items:

- The levels of security required for different data and metadata and environments, and how these are supported.
- The IT security system, employees with roles related to security (e.g. security officers), and any risk analysis approach in use.
- Measures in place to protect the facility. How the premises where digital objects are held are secured.
- Any security-specific standards the repository references or complies with.
- Any authentication and authorization procedures employed to securely manage access to systems in use.

Responses should not cover Storage and Integrity (R14) measures or the wider Technical Infrastructure (R15).

Responses and evidence should demonstrate that the applicant understands all risks to the digital and physical environment applicable to the service provided to the Designated Community. Mechanisms must be in place to prevent, detect, and respond to a security incident.

Authentication and authorization procedures must be sufficient to guarantee the security of data and metadata at each stage of the workflow (e.g. by requiring two-factor authentication for sensitive information).

Evidence should include which security policies are in place to govern the security of all systems, including network security, intrusion checks, physical facility security, and passwords.

If the response to Legal and Ethical (R04) includes references to holding sensitive digital objects then the response here should be explicit about the additional measures taken to protect this data and metadata.

Applicant Feedback

We welcome feedback on the CoreTrustSeal Requirements and the Certification procedure.

Response