DSA–WDS Partnership Working Group Catalogue of Common Requirements

Introduction

Importance of Certification

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that the data they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.

Core Certification and its Benefits

Nowadays certification standards are available at different levels, from a core level to extended and formal levels. Even at the core level, certification offers many benefits to a repository and its stakeholders.

Core certification involves a minimally intensive process whereby digital repositories supply evidence that they are sustainable and trustworthy. A repository first conducts an internal self-assessment, which is then reviewed by community peers. Such assessments help data communities—producers, repositories, and consumers—to improve the quality and transparency of their processes, and to increase awareness of and compliance with established standards. This community approach guarantees an inclusive atmosphere in which the candidate repository and the reviewers closely interact.

In addition to external benefits, such as building stakeholder confidence, enhancing the reputation of the repository, and demonstrating that the repository is following good practices, core certification provides a number of internal benefits to a repository. Specifically, core certification offers a benchmark for comparison and helps to determine the strengths and weaknesses of a repository.

Completing a self-assessment is very useful even if a repository does not wish to apply for core certification since it enables an appraisal of the repository's internal procedures, which can be examined with respect to relevant criteria and updated where necessary. The current status of the repository is therefore made apparent, and can also serve for prospective accreditation. By submitting the application for review, the repository's procedures and documentation are additionally evaluated by external professionals, taking into account the specific aims and context; thus, the repository gains independent insights on how it may evolve and mature to further increase its trustworthiness. Finally, core certification offers a solid foundation for the repository to apply for a higher-level certification in the future.

With these benefits in mind, we encourage repositories to explore core certification, including the Catalogue of Common Requirements presented here.

Background & General Guidance

This Catalogue of Common Requirements was developed by the DSA–WDS Partnership Working Group on Repository Audit and Certification, a Working Group (WG) of the Research Data Alliance[1]. The goal of the effort was to create a set of harmonized Common Requirements for certification of repositories at the core level, drawing from criteria already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS). An additional goal of the project was to develop Common Procedures to be implemented by both DSA and ICSU-WDS. Ultimately, the DSA and ICSU-WDS plan to collaborate on a global framework for repository certification that moves from the core to the extended (nestor-Seal DIN 31644), to the formal (ISO 16363) level.

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

Each Requirement in the Catalogue is accompanied by guidance text to assist applicants in providing sufficient evidence that their repositories meet the Requirement, outlining the types of information that a reviewer will expect in order to perform an objective assessment. Furthermore, the applicant must indicate a compliance level for each of the Requirements:

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

Compliance levels provide a useful part of the self-assessment process, but all applicants will be judged against statements supported by appropriate evidence; not against self-assessed compliance levels. In this regard, if the applicant believes a Requirement is not applicable, the reason for this must be documented in detail. Note also that compliance levels 1 and 2 can be valid for internal self-assessments, while certification may be granted if some guidelines are considered to be at level 3—in the implementation phase—since the Requirements include an assumption of a repository's continuous improvement.

Reponses must be in English. Although attempts will be made to match reviewers to applicants in terms of language and discipline, this is not always possible. If evidence is in another language, an English summary must be provided in the self-assessment.

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

Repositories are required to be reassessed under the DSA–WDS Common Certification standard every three years. It is recognized that while basic systems and capabilities evolve continuously according to technology and user needs, they may not undergo major changes in this timeframe. However, the Trusted Digital Repository ISO standard (ISO 16363) has a five-year review cycle, and a shorter period is considered necessary for a core trust standard to allow for possible modifications and corrections. Hence, an organization with well-managed records and business processes should reasonably expect to be able to submit an application with only minimal revisions after three years, unless the Requirements themselves have been updated within the intervening period.

[1] See the Case Statement for the RDA Working Group and the full set of Working Group members. Members of the Working Group that created this document included the following individuals representing the Data Seal of Approval and ICSU World Data System: Michael Diepenbroek, Ingrid Dillo, Rorie Edmunds, Francoise Genova, Li Guoqing, Wim Hugo, Hervé L'Hours, Jean-Bernard Minster, Mustapha Mokrane, Lesley Rickards (Co-Chair), Paul Trilsbeek, Mary Vardigan (Co-Chair).

Governance

DSA and ICSU-WDS agree to ongoing collaboration to ensure effective stewardship of the Common Requirements and Common Procedures. To that end, the DSA Board and WDS Scientific Committee will meet periodically to review and update the Common Requirements and Common Procedures and to issue new versions.

Glossary of Terms

Please refer to Appendix A for a glossary of terms.

Common Requirements

Background Information

Context

R0. Please provide context for your repository.

– *Repository Type*. Select all relevant types from[2]:

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

Comments

– *Brief Description of the Repository's Designated Community*

– *Level of Curation Performed*. Select all relevant types from[3]:

A. **Content distributed as deposited**
B. **Basic curation – e.g., brief checking, addition of basic metadata or documentation**
C. **Enhanced curation – e.g., conversion to new formats, enhancement of documentation**
D. **Data-level curation – as in C above, but with additional editing of deposited data for accuracy**

Comments

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

– *Other Relevant Information*

Response

Guidance:
To assess a repository, reviewers need some information about the repository to set it in context. Please select from among the options and provide details for the items that appear in the Context requirement.

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

(2) Repository's Designated Community. This item will be useful in assessing how the repository interacts and communicates with its target community. Please make sure that the response is specific—for example, 'quantitative social science researchers and instructors'.

(3) Level of Curation. This item is intended to elicit whether the repository distributes its content to data consumers without any changes, or whether the repository adds value by enhancing the content in some

[2] Examining the literature, the WG used this paper by Armbruster & Romary as a starting point to generate a list of repository categories as a controlled ontology. The list was then refined by the WG in collaboration with the RDA–WDS Publishing Data Cost Recovery for Data Centres. A free-text 'Other' option has also been included as a mechanism to evolve this list in the future.

[3] Likewise, the WG generated this controlled ontology in collaboration with the RDA–WDS Publishing Data Cost Recovery for Data Centres. The list may also evolve over time according to the responses given to R11 concerning Data Quality, where curation can be explained at length by an applicant.

way All levels of curation assume initial deposits are retained unchanged and that edits are only made on copies of those originals. Annotations/edits must fall within the terms of the licence agreed with the data producer and be clearly within the skillset of those undertaking the curation. Thus, the repository will be expected to demonstrate that any such annotations/edits are undertaken and documented by appropriate experts and that the integrity of all original copies is maintained. Knowing this will help reviewers in assessing other certification requirements. Further details can be added that would help to understand the levels of curation you undertake.

(4) Outsource Partners. Please provide a list of Outsource Partners that your organization works with, describing the nature of the relationship (organizational, contractual, etc.), and whether the Partner has undertaken any Trusted Digital Repository assessment. Such relationships may include, but are not limited to: any services provided by an institution you are part of, storage provided by others as part of multicopy redundancy, or membership in organizations that may undertake stewardship of your data collection when a business continuity issue arises. Moreover, please list the certification requirements for which the Partner provides all, or part of, the relevant functionality/service, including any contracts or Service Level Agreements in place. Because outsourcing will almost always be partial, you will still need to provide appropriate evidence for certification requirements that are not outsourced and for the parts of the data lifecycle that you control. Qualifications/certifications—including, but not limited to, the DSA or WDS certifications—are preferred for outsource partners. However, it is not a necessity for them to be certified. We understand that this can be a complex area to define and describe, but such details are essential to ensure a comprehensive review process.

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

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

Organizational Infrastructure

I. Mission/Scope

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

Compliance Level 

Response

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

For this Requirement, please describe:

- Your organization's mission in preserving and providing access to data, and include links to explicit statements of this mission.
- The level of approval within the organization that such a mission statement has received (e.g., approved public statement, roles mandated by funders, policy statement signed off by governing board).

II. Licenses

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

Compliance Level 

Response

Guidance:
Repositories must maintain all applicable licenses covering data access and use, communicate about them with users, and monitor compliance. This Requirement relates to the access regulations and applicable licenses set by the data repository itself, as well as any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information. Reviewers will be seeking evidence that the repository has sufficient controls in place according to the access criteria of their data holdings, as well as evidence that any relevant licences or processes are well managed.

For this Requirement, please describe:

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

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

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

III. Continuity of access

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

Compliance Level 

Response

Guidance:
This Requirement covers the measures in place to ensure access to, and availability of, data holdings, both currently and in the future. Reviewers are seeking evidence that preparations are in place to address the risks inherent in changing circumstances.

For this Requirement, please describe:

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

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

IV. Confidentiality/Ethics

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

Compliance Level 

Response

Guidance:
Adherence to ethical norms is critical to responsible science. Disclosure risk—for example, the risk that an individual who participated in a survey can be identified or that the precise location of an endangered species can be pinpointed—is a concern that many repositories must address. Evidence sought is concerned with not only having good practices for data with disclosure risks, but also the necessity to maintain the trust of those agreeing to have personal/sensitive data stored in the repository.

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

- How does the repository comply with applicable disciplinary norms?
- Does the repository request confirmation that data collection or creation was carried out in accordance with legal and ethical criteria prevailing in the data producer's geographical location or discipline (e.g., Ethical Review Committee/Institutional Review Board or Data Protection legislation)?
- Are special procedures applied to manage data with disclosure risk?
- Are data with disclosure risk stored appropriately to limit access?

- Are data with disclosure risk distributed under appropriate conditions?
- Are procedures in place to review disclosure risk in data, and to take the necessary steps to either anonymize files or to provide access in a secure way?
- Are staff trained in the management of data with disclosure risk?
- Are there measures in place if conditions are not complied with?
- Does the repository provide guidance in the responsible use of disclosive, or potentially disclosive data?

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

V. Organizational infrastructure

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

Compliance Level

Response

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

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

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

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

VI. Expert guidance

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

Compliance Level 

Response

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

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

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

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

Digital Object Management

VII. Data integrity and authenticity

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

Compliance Level: 

Response

Guidance:
The repository should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access.

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

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

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

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

Evidence of authenticity management should relate to the follow questions:

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

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

VIII. Appraisal

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

Compliance Level 

Response

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

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

- Does the repository use a collection development policy to guide the selection of data for archiving?
- Does the repository have quality control checks to ensure the completeness and understandability of data deposited? If so, please provide references to quality control standards and reporting mechanisms accepted by the relevant community of practice, and include details of how any issues are resolved (e.g., are the data returned to the data provider for rectification, fixed by the repository, noted by quality flags in the data file, and/or included in the accompanying metadata?)
- Does the repository have procedures in place to determine that the metadata required to interpret and use the data are provided?
- What is the repository's approach if the metadata provided are insufficient for long-term preservation?
- Does the repository publish a list of preferred formats?
- Are quality control checks in place to ensure that data producers adhere to the preferred formats?
- What is the approach towards data that are deposited in non-preferred formats?

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

IX. Documented storage procedures

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

Compliance Level 

Response

Guidance:
Repositories need to store data and metadata from the point of deposit, through the ingest process, to the point of access. Repositories with a preservation remit must also offer 'archival storage' in OAIS terms.

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

- How are relevant processes and procedures documented and managed?
- What levels of security are required, and how are these supported?
- How is data storage addressed by the preservation policy?
- Does the repository have a strategy for backup/multiple copies? If so, what is it?
- Are data recovery provisions in place? What are they?
- Are risk management techniques used to inform the strategy?
- What checks are in place to ensure consistency across archival copies?

- How is deterioration of storage media handled and monitored?

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

X. Preservation plan

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

Compliance Level 

Response

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

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

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

XI. Data quality

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

Compliance Level 

Response

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

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

For this Requirement, please describe:

- The approach to data and metadata quality taken by the repository.
- Any automated assessment of metadata adherence to relevant schema.
- The ability of the Designated Community to comment on, and/or rate data and metadata.
- Whether citations to related works or links to citation indices are provided.

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

XII. Workflows

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

Compliance Level 

Response

Guidance:
To ensure the consistency of practices across datasets and services and to avoid ad hoc and reactive activities, archival workflows should be documented, and provisions for managed change should be in place. The procedure should be adapted to the repository mission and activities, and procedural documentation for archiving data should be clear.

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

- Workflows/business process descriptions.
- Clear communication to depositors and users about handling of data.
- Levels of security and impact on workflows (guarding privacy of subjects, etc.).
- Qualitative and quantitative checking of outputs.
- Appraisal and selection of data.
- Approaches towards data that do not fall within the mission/collection profile.
- The types of data managed and any impact on workflow.
- Decision handling within the workflows (e.g., archival data transformation).
- Change management of workflows.

This Requirement confirms that all workflows are documented. Evidence of such workflows may have been provided as part of other task-specific Requirements, such as for ingest in R8 (Appraisal), storage procedures in R9 (Documented storage procedures), security arrangements in R16 (Security), and confidentiality in R4 (Confidentiality/Ethics).

XIII. Data discovery and identification

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

Compliance Level 

Response

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

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

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

XIV. Data reuse

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

Compliance Level 

Response

Guidance:
Repositories must ensure that data can be understood and used effectively into the future despite changes in technology. This Requirement evaluates the measures taken to ensure that data are reusable.

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

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

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

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

Technology

XV. Technical infrastructure

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

Compliance Level 

Response

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

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

- What standards does the repository use for reference? Are these international and/or community standards (e.g., Spatial Data Infrastructure (SDI) standards, OGC, W3C, or ISO 19115)? How often are these reviewed?
- How are the standards implemented? Are there any significant deviations from the standard? If so, please explain.
- Does the repository have a plan for infrastructure development? If so, what is it?
- Is a software inventory maintained and is system documentation available?
- Is community-supported software in use? Please describe.
- For real-time to near real-time data streams, is the provision of around-the-clock connectivity to public and private networks at a bandwidth that is sufficient to meet the global and/or regional responsibilities of the repository?

XVI. Security

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

Compliance Level 

Response

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

For this Requirement, please describe:

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

This Requirement describes some of the aspects generally covered by others—for example, R12

(Workflows)—and is supplementary to R9 (Documented storage procedures).

Applicant Feedback

Comments/feedback

The DSA–WDS Catalogue of Common Requirements is not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response

Appendix A: Glossary of Terms

(* Taken from OAIS[4])

Access Rights Information*: The information that identifies the access restrictions pertaining to the Content Information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the repository) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures.

Archive*: An organization that intends to preserve information for access and use by a Designated Community.

Authenticity*: The degree to which a person (or system) regards an object as what it is purported to be. Authenticity is judged on the basis of evidence.

Consumer*: The role played by those persons, or client systems, who interact with repository services to find preserved information of interest and to access that information in detail. This can include other repositories, as well as internal repository persons or systems.

Data*: A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.

Designated Community*: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time.

Digital Migration*: The transfer of digital information, while intending to preserve it, within the repository. It is distinguished from transfers in general by three attributes: a focus on the preservation of the full information content that needs preservation; a perspective that the new archival implementation of the information is a replacement for the old; and an understanding that full control and responsibility over all aspects of the transfer resides with the repository.

Digital Object*: An object composed of a set of bit sequences.

Domain or subject-based repository: Specializes in a specific research field or data type.

Ingest: The process of entering data and associated metadata into a data repository.

Integrity: Internal consistency or lack of corruption of digital objects. Integrity can be compromised by hardware errors even when digital objects are not touched, or by software or human errors when they are transferred or processed.

Institutional repository: Generic, multi-subject repository serving a research performing institution.

Long Term*: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in a repository. This period extends into the indefinite future.

Long-term Preservation*: The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term.

National repository system, including governmental: Multidisciplinary, national infrastructure. Has a legal mandate for certain (public or governmental) data types.

Open Archival Information System (OAIS)*: An Archive, consisting of an organization, which may be part of a larger organization, of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities that allows an OAIS Archive to be distinguished from other uses of the term 'Archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open

[4] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Recommended Practice -- CCSDS 650.0-M-2. Magenta Book, June 2012.http://public.ccsds.org/publications/archive/650x0m2.pdf

forums, and it does not imply that access to the Archive is unrestricted.

Preferred Formats: Formats that a repository can reasonably assure will remain readable and usable. Typically, these are the de facto standards employed by a particular discipline.

Producer*: The role played by those persons or client systems that provide the information to be preserved. This can include other repositories or internal repository persons or systems.

Provenance Information*: The information that documents the history of the Content Information. This information tells the origin or source of the Content Information, any changes that may have taken place since it was originated, and who has had custody of it since it was originated. The Archive is responsible for creating and preserving Provenance Information from the point of Ingest; however, earlier Provenance Information should be provided by the Producer. Provenance Information adds to the evidence to support Authenticity.

Publication repository: Generic, multidisciplinary repository, focussing on data linked to publications.

Reference Model*: A framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment. A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist.

Research project repository: Capture research results that require a deposit mandate by a funder or organization as a 'record of science'. Often tied to a specific (multi)disciplinary project.

Reuse: The use of data collected for one purpose to study a new problem or to verify the conclusions of the data producer.

Succession Plan*: The plan of how and when the management, ownership and/or control of the repository holdings will be transferred to a subsequent repository in order to ensure the continued effective preservation of those holdings.