



# DSA–WDS Partnership Working Group Results of Common Certification Testbed

---

*The following Members of the Working Group<sup>1</sup> contributed to the creation of this document: Michael Diepenbroek, Ingrid Dillo<sup>2</sup>, Rorie Edmunds<sup>2</sup>, Françoise Genova, Hervé L'Hours<sup>2</sup>, Wim Hugo, Varsha Khodiyar, Jean-Bernard Minster, Mustapha Mokrane, Eleni Panagou, Lesley Rickards (Co-chair), Paul Trilsbeek, Mary Vardigan (Co-chair)*

## Contents

- [Acknowledgements](#)
- [Introduction](#)
- [Purpose & Methodology](#)
  - [Testbed Volunteers](#)
- [General Outcomes](#)
- [Conclusions](#)
- [Common Requirements](#)
  - [Background Information](#)
    - [Context](#)
  - [Organizational Infrastructure](#)
    - [I. Mission/Scope](#)
    - [II. Licenses](#)
    - [III. Continuity of access](#)
    - [IV. Confidentiality/Ethics](#)
    - [V. Organizational infrastructure](#)
    - [VI. Expert guidance](#)
  - [Digital Object Management](#)
    - [VII. Data integrity and authenticity](#)
    - [VIII. Appraisal](#)
    - [IX. Documented storage procedures](#)
    - [X. Preservation plan](#)
    - [XI. Data quality](#)
    - [XII. Workflows](#)
    - [XIII. Data discovery and identification](#)
    - [XIV. Data reuse](#)
  - [Technology](#)
    - [XV. Technical infrastructure](#)
    - [XVI. Security](#)
  - [Additional Information & Applicant Feedback](#)
    - [XVII. Additional information](#)
    - [XVIII. Applicant feedback](#)
- [Appendix A: Glossary of Terms \(Taken from OAIS\)](#)
  - [Terms Not Present in the OAIS Glossary](#)

---

<sup>1</sup> See [here](#) for the full set of Working Group Members.

<sup>2</sup> Principal author.



## Acknowledgments

The WG would like to begin by expressing its huge thanks to all of the volunteers for the work they put into completing their testbed responses; they were extremely helpful in bringing relevant issues to the attention of the WG. In particular, we acknowledge the following organizations, who agreed to be named within this report so that their time and effort could be appropriately recognized:

- Flanders Marine Institute
- GESIS Data Archive for the Social Sciences
- International Service of Geomagnetic Indices
- WDC - Geomagnetism, Kyoto
- WDC - Meteorology, Asheville

## Introduction

In its RDA case statement, one of the deliverables stated by the Repository Audit and Certification DSA–WDS Partnership Working Group (WG) was

*Developing a common testbed, and its surrounding organizational framework, for peer review and certification... [to] provide practical insight into the proposed common WDS–DSA catalogue and review process, thus enabling iterative improvements to those procedures. The testbed will be driven by the DSA Board and the WDS Scientific Committee. A pool of reviewers will be set up to test the common procedures developed.*

The deliverable was thus a collaborative effort to establish a common testbed and board of reviewers towards the practical implementation of the Catalogue of Common Requirements developed by the DSA–WDS Partnership WG.

Since the process itself cannot be shared as a tangible output, the following report is the embodiment of the deliverable, and describes the purpose and methodology of the testbed, the general consequences for the Common Requirements and Procedures from an evaluation by the WG of the testbed results, as well as a more in depth look at comments received and highlighted issues for each of the Common Requirements.

## Purpose & Methodology

The main reasons behind performing the testbed exercise were to

1. Respond to critiques and to amend the Requirements if needed.
2. Examine the types of application responses that might be received. Assumptions made about comprehension of questions and content of likely responses during the design process can never be sufficient. Only through user testing can we evaluate real-world experience of the process.
3. Make sure of the bilateral nature of the Common Certification by determining whether an organization can be considered suitable for certification by DSA, ICSU-WDS, or both through completion of the Common Requirements.

It would thus act as a trial run of the whole process from beginning to end, as if assigning the DSA–WDS certification.

To conduct the testbed, the WG created a Google Form following the Catalogue of Common Requirements to capture tester's responses; the spreadsheet behind the Form enabling the WG to easily evaluate the volunteers' responses individually or jointly as a group to compare across.

DSA and WDS then approached Members within their respective communities to 'apply' against the Common Requirements; specifically, the exercise involved volunteers from the two organizations responding to the Common Requirements and having their responses read by all members of the WG from the viewpoints of

- A pure comparison. Do the responses for each Requirement contain the expected types of information and is the evidence reasonably consistent?
- A review. Would the repository be awarded a certification based on the responses?

In addition, the WG decided to identify two reviewers for each application—one from DSA and one from WDS—to conduct a more in depth, formal assessment, with the consequence of potentially granting the test



organizations both WDS membership and the DSA Seal. This formal review was yet to commence at the time of writing and its results are considered outside the scope of the current work; however, it is mentioned here to highlight the progress that the organizations are making towards a joint board and pool of reviewers for the nascent DSA–WDS certification standard.

## Testbed Volunteers

Two testers were approached by the DSA Board, neither of which were native English speakers. The WG was especially careful to try and ensure that those not having English as a first language be taken into consideration when creating the Common Requirements, and it was thus keen to have at least one non-native English speaker answer the Requirements.

ICSU-WDS identified four WDS Member Organizations to approach for the testbed, since this was considered more fully representative of its Members, and therefore beneficial from a WG standpoint. These volunteers represented: an organization under the former World Data Centre system, an organization under the former Federation of Astronomical and Geophysical data-analysis Services, an organization that was not part of either of these regimes, and an organization outside of Europe/North America. The WG felt it was particularly valuable to assess the evidence supplied by a data service such that this might be added to the controlled vocabulary used in the Common Requirements.

## General Outcomes

In the pages that follow, we provide a version of the Common Requirements in which the WG has examined in turn the comments and evidence received for each Requirement. In all cases, the WG has attempted to provide appropriate responses/actions that lead back to revisions for incorporation into the Common Requirements and Procedures documents. These documents will be taken forward by DSA and WDS as the first, official version of the common Core Certification Standard for Trustworthy Data Repositories.

It is also clear from the testbed that whilst amended versions of the current guidance might be kept alongside the Requirements, a separate, extended set of *Additional Guidance for Applicants* is needed to give further, highly detailed information that is either too long or too complex to be included within the Requirements document. Such Additional Guidance may be created by the future joint Certification Authority (when it exists) and would likely include some/all of the following:

- Advice on how to develop and manage minimal, but best practice, public statements that can be used to support the evidence within the Requirements. The WG is aware that not all information can be publically available due to security issues, and this will be accounted for in the process. That website links can change over time must be further discussed by DSA and WDS since this, by implication, affects the trustworthiness of a repository.
- Further definition of the range of organizational expertise that must be brought together to respond to the Requirements.
- More explicit explanation about the different types of metadata will be attempted and a diagram produced to address this point. Metadata is mentioned extensively throughout the Common Requirements and can seem rather repetitive, but this is unavoidable since most aspects of data stewardship require metadata.
- Answering the rightful concern for the curation levels listed in R0 (Context), that a repository may often not be able to judge about data accuracy. The WG has proposed that a brief description be written to clarify the subtle differences between the task of evaluating technical quality in terms of good data curation practice and the issue of research quality. The latter is beyond the scope of core certification of a trustworthy digital repository.
- The different approaches to and expectations of appraisal between repositories that perform additional data services and that may handle ‘early versions’ of collected data, and those with a pure repository function. Such distinctions among data repositories is something that the WG is highly cognizant of, and will be explored in future discussions between DSA and WDS (possibly in collaboration with the RDA/WDS Certification of Digital Repositories IG).
- That the level of formal process documentation remains extremely variable across repositories. Both WDS and DSA know this from past experience, and are keen to see more standardized, minimum expectations, but we understand that this is a challenge with such a wide variety of organizations.



## Conclusions

The testbed has presented WDS and DSA with a firm basis for moving forwards together. As a result of its findings, we will seek to provide both amended and additional guidance, and to offer a process that fully supports applicants. Explicitly, immediate tasks will be to:

- Offer tools and other supporting materials that meet the confidentiality needs of the review process, as well as the public provision of evidence.
- Clarify the cross-disciplinary nature of the Requirements, with particular regard to explaining ethical considerations that may apply to only a subset of scientific data.
- Further define the descriptions of an organization's scope to enable inclusion of teams and partnerships, in contrast to more formal provisions of outsourcing services.
- Make reviewers aware when assessing applicants that continuity of funding is seldom guaranteed, and this must be balanced with the need for stability.
- Consider whether R17 (additional information) should instead form part of the opening contextual section of the Requirements (i.e., R0).
- Further expand upon and explain data management and OAIS repository concepts that might not be familiar to all applicants.

Producing the Common Requirements and Procedures is an evolving process, and the WG did not reasonably expect to have everything perfectly in place first time around. Now that the WG has gone through the feedback received through the testbed, it will revise the Common Requirements accordingly, and then the finalized RDA product will be adopted by DSA and ICSU-WDS. From thereon, the two organizations will continue to work together to improve the Requirements and Procedures, and at regular intervals will examine whether they need to be fine-tuned depending on developments within the landscape. The WG is very happy that it was given the opportunity to do this work within the Research Data Alliance; the idea of creating a Common Certification began under its auspices, and DSA and ICSU-WDS were inspired by RDA to realize this core certification standard together. The two organizations must now take the outputs of this endeavour and actually use them to stimulate more accreditations, increase impact on the community, and continue the development of a common framework for certification and a service of trustworthy data repositories.



## Common Requirements

### Testbed Results – Common Requirements Process

#### Selected General Comments from Testers

- 'Clear and understandable'.
- 'The structure seems to be more suitable'.
- 'Stronger emphasis on documented procedures and plans'.
- 'In total it is easier to meet the common requirements but there are also some that need a more detailed description and evidence'.

#### WG Comments on Issues Raised

- Duplication across questions.  
*While unavoidable to some extent, repetition should be minimized, and the material will be reviewed with this in mind.*
- Repetition of requests for metadata information.  
*We may consider either clarifying the types of metadata required (always a risk as administrative, preservation, descriptive, and other metadata overlap and are subject to local context) or reviewing the use of the term 'metadata' and seeking clearer alternatives.*
- Experience of using the Google Form and associated PDF form, including presentation of an initial overview of the Requirements, the need to share with colleagues, providing formatting, and better navigation and submission options.  
*These will be taken forward in the future development of supporting tools for the Common Certification process.*
- Maximum response length.  
*For existing DSA applicants, the evidence is public and therefore provides a guide to likely length. However, those submissions vary greatly (e.g., outsourcing may require significant additional context). We will consider suggesting—but not necessarily technically enforcing—a maximum length once further tests are undertaken.*
- Mandatory topics and their weighting.  
*Further clarification will be added to the introduction to indicate that all Requirements are mandatory and there is no 'weighting' across them; each is a standalone item.*



## Background Information

### Context

**R0. Please provide context for your repository.**

– **Repository Type.** Select from<sup>3</sup>:

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

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

– **Level of Curation Performed.** Select from<sup>4</sup>:

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

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

Response

Guidance:

To assess a repository, reviewers need some information about the repository to set it in context. Please select from among the options for the four contextual 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.

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

**(3) Level of Curation.** This item is intended to elicit whether the repository distributes its content to data consumers without any changes, or whether the repository adds value by enhancing the content in some way. Knowing this will help reviewers in assessing other certification requirements.

**(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 Partners 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.

<sup>3</sup> 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.

<sup>4</sup> 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.



## Testbed Results

### WG Comments on Issues Raised

- Space to add further contextual information for the multiple choice criteria.  
*We will provide the option to add comments under the multiple choice criteria since this will support their evolution in the future.*
- Descriptions of the repository types.  
*We will include definitions in the glossary for the terms in the multiple choice criteria.*
- For enhanced curation, is 'creation of new formats' during the ingest process or the logical migration process?  
*The confusion here stems from the usage of 'creation of new formats'. On reflection, this will be changed to 'conversion to new formats', which we hope should make clearer its intended meaning.*
- Clarity on 'Outsource Partners'. Checking whether an Outsource Partner has undertaken a Trusted Digital Repository assessment is not straightforward, in general.  
*We will provide additional clarification on outsourcing and the associated evidence required; noting that this is a complex area in which organizations may rely on internal institutional partners or have a full contractual relationship with a third party. All information supplied by an organization supports the understanding of its context, and in this regard, Partners with relevant qualifications/certifications—including, but not limited to the DSA or WDS certifications—are preferred, but it is not a necessity for them to be certified.*
- Editing of data contradicts R7 (Data integrity and authenticity). A repository may not be able to judge on accuracy, and data should be edited only at the request of the data provider/intellectual rights owner.  
*We will add the following above the 'levels of curation' definitions to clarify that (1) there is no contradiction between requiring integrity and undertaking curation that may amend the dissemination copies of the data, and (2) only appropriately skilled curators should undertake data-level curation: 'All levels of curation assume initial deposits are retained unchanged and that amendments are only made on copies of those originals. Such amendments must fall within the terms of the licence agreed with the data producer and be clearly within the skillset of those undertaking the curation'. We may also consider replacing 'editing' with 'annotating/editing'.*



## 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:

- Explicit statements of this role within the organization’s mission and provide links.
- 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).

**Testbed Results**

**WG Comments on Issues Raised**

- Clarify ‘this role’ in ‘Explicit statements of this role’.  
*We will consider changing to ‘Please provide statements (with links) that make explicit your organization’s mission in providing access to and preservation of data’.*  
*Furthermore, we will seek to provide additional support on describing the organizational scope (in terms of which ‘entity’ of ‘data collections’ or ‘services’ within an entity is seeking certification) and how this contrasts with cases of outsourcing.*





**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.

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.

This Requirement must be read in conjunction with R4 (Confidentiality/Ethics) to the extent that ethical and privacy provisions impact on the licenses.

**Testbed Results**

**WG Comments on Issues Raised**

- Clarify ‘monitors compliance’; WDS Members must have Open Data, in principle.  
*DSA has many members who champion the concept of Open Data, but for legal or ethical reasons must control access to some/all of their data collections. ‘Open’ itself does not mean completely public and without conditions imposed on the user, but if this is the case for a whole data collection, the applicant can simply say so. In many other cases, data could be ‘open’, but the applicant might describe conditions of use such as requiring attribution or agreement to make any secondary analysis available under the same ‘open’ conditions.*  
*Reviewers will be seeking evidence that the applicant has sufficient controls in place according to the access criteria of their data collection, as well as evidence that any relevant licences or processes are well managed.*



### 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.

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.

#### **Testbed Results**

##### **WG Comments on Issues Raised**

- Overlap between Requirements R3 (Continuity of Access) and R15 (Technical Infrastructure) / R16 (Security).

*R15 covers evidence that the technical infrastructure is fit for its primary intended purpose, while R16 focusses on the security (including nontechnical) of operations. In contrast, this Requirement covers full business continuity of the preservation and access functions. Reviewers are seeking evidence of a mature organization prepared to address the risks inherent in changing circumstances.*

##### **Actions Based on Testbed Responses**

- We may need to further refine the guidance here to ensure that both organizational continuity and access risk in terms of preservation strategies (emulation, format migration, etc.) are more explicit.



**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.

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).

**Testbed Results**

**WG Comments on Issues Raised**

- Intimacy of questions, considering that some repositories might be in 'competition' with one another. Similar questions to R4 might be asked of DSA and WDS.  
*This is a reasonable statement, and provision is made for applicants who wish commercially or security sensitive parts of their evidence to remain confidential. But in the case of R4, the evidence sought is not simply about having good practices for data with disclosure risks, it also concerns the necessity to maintain the trust of those agreeing to have personal/sensitive data stored in the repository.*  
*That DSA and WDS might be asked similar questions is an excellent point, and one that is important to both organizations, who want to function as open, trusted 'nodes' in the scientific infrastructure.*

**Actions Based on Testbed Responses**

- We might provide further guidance on how best to communicate that there are no ethical issues due to the nature of the data holdings.



### 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.

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.

**Testbed Results**

**Actions Based on Testbed Responses**

- A note will be added that clearer descriptions of processes undertaken, and the skills necessary to perform them, would be helpful from applicants; but while this is relevant, it is well beyond the sphere 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?

**Testbed Results**

**Actions Based on Testbed Responses**

- There was a distinct overlap between R5 (Organizational infrastructure) and R6, where in both cases, respondents referred to staff skills. This would benefit from some further consideration and clarification in the Requirements.



## 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.


#### **Testbed Results**

Test responses were generally good and consistent.



## 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.

### **Testbed Results**

Test responses were generally good and consistent.



### 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.

**Testbed Results**

Test responses were generally good and consistent.





**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 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?
- Is a preservation plan in place?
- 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?

**Testbed Results**

**WG Comments on Issues Raised**

- A repository's rights to copy, transform, and store items is not clear. Can a repository preserve data, but not copy them? Is this transformation of only the format, the content, or...?  
*One intention of the guidance is to drive internal communications within applicant organizations. We are aware of repositories whose older deposits were licenced with little regard to the issue of file format risk, which assumed bit-level storage only, and either explicitly or implicitly made format migration impossible under strict interpretation of the licence. There have been more recent cases where (meta)data transformation rights were 'assumed' but not explicit, which then presented legal barriers or required expensive renegotiation of licences before collections could be made available via new technologies.*

**Actions Based on Testbed Responses**

- We may consider further explicit guidance describing possible preservation methods, including emulation, formats, migration, and so on.



## 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. 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.

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).

### Testbed Results

#### WG Comments on Issues Raised

- Overlap between R11 and R14 (Data reuse), both of which ask if adequate metadata is provided to the user.

*We use 'metadata' throughout the Common Requirements simply as a shorthand for 'information that describes the data'. It is used in preference to 'information' since it tends to suggest something more structured, standard compliant, and (ideally) machine actionable. The topic for R11 is data quality, whereas R14 is concerned with making data understandable to a designated community of users. Quality assessment becomes increasingly relevant when research is undertaken across disciplines, where researchers may not have the personal experience to make such evaluations of quality from the data alone.*

- Data provided by WDS Members must be useful for promoting new science, and therefore it is unreasonable to request that they be very strict in keeping data quality high.  
*The key term here is 'sufficient'. We do not (and could not) seek to set a standard barrier for technical quality. An understanding that data, or associated metadata, may have quality issues is important to their research value, but does not preclude their use in science if the level of 'quality' is appropriately documented for the repository to decide whether it meets their collection criteria and for the data user to decide whether it is appropriate for use in their research.*



## 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).

### Testbed Results

Test responses were generally good and consistent.



### 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?

#### **Testbed Results**


##### **WG Comments on Issues Raised**

- It is unclear whether appropriate (internationally agreed) standards in the guidance refers to metadata standards or otherwise.  
*It refers to all standards that enable the repository to be interoperable with anyone wishing to access/harvest its (meta)data or vice-versa. In this regard, it includes metadata standards, as well as ontologies and so on. We will add text elaborating on this.*



#### 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?

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

#### Testbed Results

##### WG Comments on Issues Raised

- No distinction between 'use' and 'reuse' of data. The latter may be unnecessary. *We can perhaps provide better support for the concept of reuse by adding a definition in the glossary, as well as incorporating the following text into the guidance: 'In environments where, for instance, secondary analysis outputs are redeposited into a repository alongside primary data, the provenance chain and associated rights issues may become increasingly complicated.'*



**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 fulfills. 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?

**Testbed Results**

**WG Comments on Issues Raised**

- We are changing our hardware. Is it a relevant answer to having a plan for infrastructure development?  
*We are seeking to provide a low barrier to meet the core certification. This obviously has challenges in the complex technical aspects of repositories. In this case, the reviewer might be seeking evidence that demonstrates forward planning of infrastructure in response to technical changes and new demands from users. Not to have such a plan may indicate that an organization cannot support preservation and access into the future.*
- Community-supported software is not in use.  
*This might be an appropriate response if context were provided. Evidence of community-supported software might indicate an organization that minimizes costs and risks through adopting resources validated as 'fit for purpose' by a wide variety of stakeholders. But in the absence of such options, evidence that in-house developed, bespoke software is the most suitable for the repository services offered is perfectly acceptable.*



### 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).

**Testbed Results**

**WG Comments on Issues Raised**

- Documentation on security aspects is not in English.  
*While we will seek to match reviewers to applicants in terms of language and discipline, this may not be possible in all cases. We therefore request that an English summary always be made available either within the application form or elsewhere.*



## Additional Information & Applicant Feedback

### XVII. Additional information

**R17. Any other relevant information you wish to provide on your repository.**

Response

Guidance:

The repository may add any extra information that is not covered in the above 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.

### **Testbed Results**

#### **Actions Based on Testbed Responses**

Elements were written here that might be better placed in R0 (Context), alongside some of the more specific questions rather than at the end of the Requirements. A textbox for additional contextual information may therefore be added to the Background Information section.





## XVIII. Applicant feedback

**R18. 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

### Testbed Results – Certification & Renewal

#### WG Comments on Issues Raised

- A document stating that a repository is trustworthy according to the DSA–WDS framework would be beneficial.  
*Recipients of the DSA or WDS certification already are given a clear logo. The DSA is linked back to the public evidence statements of the repository and is displayed at an agreed web-location (one clearly associated with the scope of the application). DSA and WDS will work together to define a means by which a trusted digital repository can be clearly recognized as having met the Common Requirements.*
- Every five years is reasonable for recertification. Basic systems and capabilities do not necessarily change every three years, even though they evolve continuously to keep up with technology and user needs.  
*After consultation, DSA and WDS have agreed to move to a recertification period of three years from two- and three–five-year cycles, respectively. The Trusted Digital Repository ISO standard (ISO 16363) has a five-year review period, but for a core trust standard, we feel that a shorter period is necessary to allow for possible revisions and corrections. An organization with well-managed records and business processes, and which has not undergone major changes in the intervening period, should reasonably expect to be able to submit an unchanged submission after three years; unless of course the Requirements themselves have been updated within that period.*
- If the reaccreditation period for WDS Members is to be reduced, the Common Catalogue—or some of its Requirements—could be reduced and overlaps avoided.  
*We will seek to clarify or eliminate perceived overlaps in response to this and future testing, but the overall length of the Requirements have already been minimized as much as possible while maintaining the desired scope. Although we are seeking to provide an entry-level ‘core’ certification, we also want to align with the more detailed certification standards.*
- Clarify when the three-year period starts for the renewal procedure. Depending on the answer, there could be a gap of just one year between a repository’s certification and the start of its next renewal or it will be reviewed once every five-years at a maximum.  
*This is a valid point, and DSA and WDS are considering the separation between releasing versions of the Requirements and the period of certification. Ideally, all accreditations would last for a period of exactly three years from the point of award, and applicants would simply update to a new version of the Requirements (if there had been any changes) at the time of their renewal.*



## Appendix A: Glossary of Terms (Taken from OAIS<sup>5</sup>)

**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.

**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.

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

**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.

**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

<sup>5</sup> 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>



education and explaining standards to a non-specialist.

**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.

### Terms Not Present in the OAIS Glossary

**Curation:** Activities required to make deposited data preservable or usable now and in the future. Depending on technological changes, curation may be required at certain points in time throughout the data lifecycle.

**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.

**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.

#### Testbed Results

##### WG Comments on Issues Raised

- Integrate the latter part of the Glossary, and mark the topics taken from OAIS, to make it more readable.

*This is a valid point and the Glossary will be updated accordingly.*