



## **Data Seal of Approval**

**Guidelines version 2010**  
**June 1, 2010**



<b>Data Seal of Approval</b>	<b>1</b>
<b>Guidelines version 2010 June 1, 2010</b>	<b>1</b>
Introduction	5
Introduction to the DSA Assessment	5
Guidelines	6
Assessment	6
1. The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.	8
Could this guideline be outsourced?	8
Required statement of compliance:	8
Applicant manual	8
Reviewer manual	8
2. The data producer provides the research data in formats recommended by the data repository.	9
Could this guideline be outsourced?	9
Required statement of compliance:	9
Applicant manual	9
Reviewer manual	9
3. The data producer provides the research data together with the metadata requested by the data repository.	10
Could this guideline be outsourced?	10
Required statement of compliance:	10
Applicant manual	10
Reviewer manual	10
4. The data repository has an explicit mission in the area of digital archiving and promulgates it.	11
Could this guideline be outsourced?	11
Required statement of compliance:	11
Applicant manual	11
Reviewer manual	11
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.	12
Could this guideline be outsourced?	12
Required statement of compliance:	12
Applicant manual	12
Reviewer manual	12
6. The data repository applies documented processes and procedures for managing data storage.	13
Could this guideline be outsourced?	13
Required statement of compliance:	13
Applicant manual	13
Reviewer manual	13
7. The data repository has a plan for long-term preservation of its digital assets.	14
Could this guideline be outsourced?	14
Required statement of compliance:	14
Applicant manual	14
Reviewer manual	14
8. Archiving takes place according to explicit work flows across the data life cycle.	15
Could this guideline be outsourced?	15
Required statement of compliance:	15
Applicant manual	15
Reviewer manual	15
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.	16
Could this guideline be outsourced?	16
Required statement of compliance:	16
Applicant manual	16

Reviewer manual	16
10. The data repository enables the users to utilize the research data and refer to them.	17
Could this guideline be outsourced?	17
Required statement of compliance:	17
Applicant manual	17
Reviewer manual	17
11. The data repository ensures the integrity of the digital objects and the metadata.	18
Could this guideline be outsourced?	18
Required statement of compliance:	18
Applicant manual	18
Reviewer manual	18
12. The data repository ensures the authenticity of the digital objects and the metadata.	19
Could this guideline be outsourced?	19
Required statement of compliance:	19
Applicant manual	19
Reviewer manual	19
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.	20
Could this guideline be outsourced?	20
Required statement of compliance:	20
Applicant manual	20
Reviewer manual	20
14. The data consumer complies with access regulations set by the data repository.	21
Could this guideline be outsourced?	21
Required statement of compliance:	21
Applicant manual	21
Reviewer manual	21
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and scientific research for the exchange and proper use of knowledge and information.	22
Could this guideline be outsourced?	22
Required statement of compliance:	22
Applicant manual	22
Reviewer manual	22
16. The data consumer respects the applicable licenses of the data repository regarding the use of the research data.	23
Could this guideline be outsourced?	23
Required statement of compliance:	23
Applicant manual	23
Reviewer manual	23

# Introduction

## Introduction to the DSA Assessment

The assessment guide connected to the guidelines is designed to help data managers who want to prepare an assessment of their repository to apply for the Data Seal of Approval (DSA). Per DSA's guideline it lists suggestions of topics for inclusion and discussion. It is neither prescriptive nor exhaustive.

Each guideline should be addressed in this assessment by a link to a publicly available statement (preferably in English) which relates to the issues noted below each guideline. If the documentation is not in English a small summary in English should be provided. If documentation is not ready for publishing a deadline needs to be mentioned which will be monitored.

Please avoid the use of abbreviations as much as possible, if abbreviations are used please provide an explanation with them.

Although the sixteen guidelines regard three stakeholders – the data producer (three guidelines), the data consumer (three guidelines) and the data archive (ten guidelines) – the data archive is seen as the main organization responsible for the repository. The data archive as an organization should take care of the overall implementation of the DSA in its own specific field.

An organization that merely provides access to data and does not do any archiving in its own repository can achieve the DSA logo provided it regularly does the following:

- Take care of the implementation of the DSA guidelines 1, 2, 3, 5, 9, 10, 11, 12, 14, 15 and 16
- Mirror the data in another Trusted Digital Repository (TDR) that has acquired the DSA logo or better. If the agency outsources to does not have the latest DSA or better, please describe why it is trustworthy. This TDR takes over the responsibility for guidelines 4, 6, 7, 8 and 13 and guarantees the long-term preservation of the data

The repository gives information as to how the guidelines relate to the repository. The responsible archivist gives his or her assessment of how the various matters have been implemented in the data repository, which main issues need to be addressed, and where visitors can find information. It should also state clearly to what other websites, if any, the DSA will apply and be explicit in their evidence.

The reviewer will check amongst others whether the documents mentioned under comments or URL are open access available to guarantee transparency. The review process will take a maximum of 2,5 months.

## Minimal requirements to get the DSA awarded in 2010.

Every year the requirements will be reviewed by the DSA Board. Over time the dot-standard will be raised according to higher quality criteria. We expect the minimum requirements in 2015 will be 4 dots for all guidelines.

If the repository does not comply to the requirements mentioned below but will do so in the near future, it is also possible to apply for a temporary Data Seal of Approval. The assessment form then needs to contain deadlines which state when the repository will be compliant. These deadlines will be monitored.

Note that this is a tool in progress that will develop further as more and more DSA assessments are performed. Please feel free to inform us of your own professional insight into the guidelines of the DSA and point our attention to the specific information you provide ([info@datasealofapproval.org](mailto:info@datasealofapproval.org)).

## Guidelines

Guideline	Minimum requirement	Guideline can be outsourced
1 The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.	3	No
2 The data producer provides the research data in formats recommended by the data repository.	3	No
3 The data producer provides the research data together with the metadata requested by the data repository.	4	No
4 The data repository has an explicit mission in the area of digital archiving and promulgates it.	4	Yes
5 The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.	4	No
6 The data repository applies documented processes and procedures for managing data storage.	4	Yes
7 The data repository has a plan for long-term preservation of its digital assets.	3	Yes
8 Archiving takes place according to explicit work flows across the data life cycle.	3	Yes
9 The data repository assumes responsibility from the data producers for access and availability of the digital objects.	4	No
10 The data repository enables the users to utilize the research data and refer to them.	2	No
11 The data repository ensures the integrity of the digital objects and the metadata.	3	No
12 The data repository ensures the authenticity of the digital objects and the metadata.	3	No
13 The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.	3	Yes
14 The data consumer complies with access regulations set by the data repository.	4	No
15 The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and scientific research for the exchange and proper use of knowledge and information.	4	No
16 The data consumer respects the applicable licenses of the data repository regarding the use of the research data.	4	No

## Assessment

Statement of Compliance	Means	Comments and/or URLs
0	N/A: Not Applicable.	Provide an explanation
1	No: We have not considered this yet.	Provide an explanation
2	Theoretical: We have a theoretical concept.	Provide a URL for the initiation document.
3	In progress: We are in the implementation phase.	Provide a URL for the supporting document.
4	Implemented: This guideline has been fully implemented for the needs of our repository.	Provide a URL for the supporting document.

### Data Seal of Approval Board

W [www.datasealofapproval.org](http://www.datasealofapproval.org)

E [info@datasealofapproval.org](mailto:info@datasealofapproval.org)



# **1. The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.**

## **Could this guideline be outsourced?**

No.

## **Required statement of compliance:**

3. In progress: We are in the implementation phase.

## **Applicant manual**

This guideline relates to the level of guidance which the repository gives to the data producer before and at the time of submission to the repository. The response should concentrate on the contribution of the repository in making compliance with this guideline possible for the data producer.

- Does the repository provide assistance to depositors in terms of describing the full package of information that should be deposited to facilitate assessment? For example, does the repository request citations based on the data? A methodology report?
- Does the repository elicit information about IRB or other official approval for data collection to confirm adherence to ethical research principles?
- Does the repository promote data sharing and reuse?
- Does the repository provide enough information to answer questions such as the following?
  - Are the research data based on work performed by the data producer (researcher or institution that makes the research available) and does the data producer have a solid reputation?
    - Data regarding, or references to, the researcher and/or research group
    - References to publications pertaining to these particular research data
  - Was data collection or digitization carried out in accordance with prevailing criteria and ethics in the research discipline?
    - Information regarding the used methods and research techniques, including the collection of those data
  - Are the research data useful for certain types of research and suitable for reuse?
    - Information regarding the manner in which the research data have obtained shape, content and structure

## **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **2. The data producer provides the research data in formats recommended by the data repository.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

3. In progress: We are in the implementation phase.

### **Applicant manual**

This guideline relates to the level of guidance which the repository gives to the data producer before and at the time of submission to the repository. The response should concentrate on the contribution of the repository in making compliance with this guideline possible for the data producer.

- Does the repository publish a list of preferred formats? Are tools used to check the compliance with official specifications of the formats? What is the repository's approach towards data that are deposited in non-preferred formats?
- Are quality control checks in place at the repository to ensure that data producers adhere to the preferred formats?
- Does the repository ask depositors to provide detailed information about their file formats and the tools and methods by which the files were created?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

### **3. The data producer provides the research data together with the metadata requested by the data repository.**

#### **Could this guideline be outsourced?**

No.

#### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

#### **Applicant manual**

This guideline relates to the level of guidance which the repository gives to the data producer before and at the time of submission to the repository. The response should concentrate on the contribution of the repository in making compliance with this guideline possible for the data producer.

- Are deposit forms which hold resource discovery metadata used?
- Are there other user-friendly ways for users to provide metadata?
- Does the repository have quality control checks to ensure that the data producer adheres to the request for metadata?
- Are there tools to create metadata at the level of files?
- Are metadata elements derived from established metadata standards, registries or conventions? If so list them, and show the level of adherence to those standards.
- Are these metadata items relevant for the data consumers?
- What is the repository's approach if the metadata provided are insufficient for long term archiving?

#### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **4. The data repository has an explicit mission in the area of digital archiving and promulgates it.**

### **Could this guideline be outsourced?**

Yes.

### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant manual**

This guideline relates to the level of authority which the repository has.

- Does the repository have a Mission Statement? Does it clearly reference a commissioning authority?
- Does the repository have a document which outlines the way in which the mission statement is implemented?
- Does the repository carry out promotional activities?
- Does the repository have succession planning in place for its digital assets? If so, please describe the plan.
- Or if applicable to which TDR have you outsourced?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.**

**Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant manual**

This guideline relates to the legal regulations which impact on the repository.

- Is the repository a legal entity? If so, please describe its legal/organizational status.
- Does the repository use model contract(s) with data producers?
- Does the repository use model contract(s) with data consumers?
- Are the repository's conditions of use published?
- Are there measures in place if the conditions are not complied with?
- Does the repository ensure knowledge of and compliance with national and international laws? How?
- 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 management of data with disclosure risk?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **6. The data repository applies documented processes and procedures for managing data storage.**

### **Could this guideline be outsourced?**

Yes.

### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant manual**

This guideline relates to the ability of the repository to manage data.

- Does the repository have a preservation policy?
- Does the repository have a strategy for backup / multiple copies? If so, please describe.
- Are data recovery provisions in place? What are they?
- Are risk management techniques used to inform the strategy?
- Are there checks on the consistency of the archive?
- What levels of security are acceptable for the repository?
- How is deterioration of storage media handled and monitored?
- Or if applicable to which TDR have you outsourced?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **7. The data repository has a plan for long-term preservation of its digital assets.**

### **Could this guideline be outsourced?**

Yes.

### **Required statement of compliance:**

3. In progress: We are in the implementation phase.

### **Applicant manual**

This guideline relates to the ability of the repository providing continued access to data.

- Are there provisions in place to take into account the future obsolescence of file formats? Please describe.
- Are there provisions in place to ensure long-term data usability? Please describe.
- Or if applicable to which TDR have you outsourced?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **8. Archiving takes place according to explicit work flows across the data life cycle.**

### **Could this guideline be outsourced?**

Yes.

### **Required statement of compliance:**

3. In progress: We are in the implementation phase.

### **Applicant manual**

This guideline relates to the levels of procedural documentation for the repository.

- Does the repository have procedural documentation for archiving data?
- If so, provide references to:
  - Workflows
  - Decision-making process for archival data transformations
  - Skills of employees
  - Types of data within the repository
  - Selection process
  - Approach towards data that do not fall within the mission
  - Guarding privacy of subjects, etc.
  - Clarity to data producers about handling of the data
- Or if applicable to which TDR have you outsourced?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant manual**

This guideline relates to the levels of responsibility which the repository takes for its data.

- Does the repository have licences / contractual agreements with data producers? Please describe.
- Does the repository enforce licences with the data producer? How?
- Does the repository have a crisis management plan? Please describe.

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **10. The data repository enables the users to utilize the research data and refer to them.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

2. Theoretical: We have a theoretical concept.

### **Applicant manual**

This guideline relates to the formats in which the repository provides its data and its identifiers.

- Are data provided in formats used by the research community? In what forms?
- Does the repository offer search facilities? Is OAI harvesting permissible? Is deep searching possible?
- Does the repository offer persistent identifiers?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **11. The data repository ensures the integrity of the digital objects and the metadata.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

3. In progress: We are in the implementation phase.

### **Applicant manual**

This guideline relates to the information contained in the digital objects and metadata and whether it is complete, whether all changes are logged and whether intermediate versions are present in the archive.

- Does the repository utilise checksums? What type? How are they monitored?
- How is the availability of data monitored?
- Does the repository deal with multiple versions of the data? If so, how? Please describe the versioning strategy.

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **12. The data repository ensures the authenticity of the digital objects and the metadata.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

3. In progress: We are in the implementation phase.

### **Applicant manual**

This guideline refers to the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

- 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, and 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?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

### **13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS.**

#### **Could this guideline be outsourced?**

Yes.

#### **Required statement of compliance:**

3. In progress: We are in the implementation phase.

#### **Applicant manual**

This guideline refers to the level of conformance with accepted standards.

- What standards does the repository use for reference?
- How is the standard implemented, and if there are significant deviations from the standard why is that the case?
- Does the repository have a plan for infrastructural development? Please describe.
- Or if applicable to which TDR have you outsourced?

#### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **14. The data consumer complies with access regulations set by the data repository.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant manual**

This guideline refers to the responsibility of the repository to create legal access agreements which relate to relevant national (and international) legislation and the levels to which the repository informs the data consumer about the access conditions of the repository.

- Does the repository use End User Licence(s) with data consumers?
- Are there any particular special requirements which the repository's holdings require?
- Are contracts provided to grant access to restricted-use (confidential) data?
- Does the repository make use of special licences, e.g., Creative Commons?
- Are there measures in place if the conditions are not complied with?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

**15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and scientific research for the exchange and proper use of knowledge and information.**

**Could this guideline be outsourced?**

No.

**Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

**Applicant manual**

This guideline refers to the responsibility of the repository to inform data users about any relevant codes of conduct.

- Does the repository need to deal with any relevant codes of conduct?
- Does the repository need to deal with codes of conduct specifically pertaining to protection of human subjects?
- What are the terms of use to which data consumers agree?
- Are institutional bodies involved?
- Are there measures in place if these codes are not complied with?
- Does the repository provide guidance in the responsible use of confidential data?

**Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.

## **16. The data consumer respects the applicable licenses of the data repository regarding the use of the research data.**

### **Could this guideline be outsourced?**

No.

### **Required statement of compliance:**

4. Implemented: This guideline has been fully implemented for the needs of our repository.

### **Applicant manual**

This guideline refers to the responsibility of the repository to inform data users regarding the applicable licences.

- Are there relevant licences in place?
- Are there measures in place if these codes are not complied with?

### **Reviewer manual**

- Does the supplied information correspond to the guideline in question? (also see the above applicant guide for this guideline)
- Are the documents mentioned under comments or URL available via open access ?
- Do you agree with the assigned symbols and are there enough to award the DSA to this guideline?
- Are abbreviations clear?
- Please add comments, not questions. As it is easier to take action on comments than questions.