



Data Seal of Approval Overview

Paul Trilsbeek
The Language Archive
Max Planck Institute for Psycholinguistics



Outline of Presentation

- Broad description of DSA and its history
- Tour through Guidelines
- Procedures and how to apply for the Seal



Key Characteristics

- Lightweight, basic certification mechanism
- 16 Guidelines for Trusted Digital Repositories
- Self-assessment, no external auditors or site visit
- Peer-reviewed process supervised by DSA Board
- DSA granted for a period of two years
- Online tool for self-assessment and review



History and Context

- Initiated by DANS in the Netherlands in 2005, first version presented in 2007
- Internationalized and handed over to international board in 2009
- First assessments done in 2010
- Now 36 repositories with the Data Seal, with another 34 in progress
- Started in social sciences and humanities but now expanding in disciplinary scope



Objectives

- Assure data producers that their data are reliably managed and preserved
- Instil confidence in funders that data in which they have invested will be available for re-use in the future
- Provide a transparent view into repositories to enable data consumers to assess them



Principles

The DSA is intended to ensure that data are:

- Available on the Internet
- Accessible
- Usable (file formats)
- Reliable
- Citable (can be referred to)



Stakeholders

Data Producer

- Responsible for the quality of the digital data (3 Guidelines)

Data Repository

- Responsible for the quality of data storage & availability (10 Guidelines)

Data Consumer

- Responsible for the quality of use of the digital data (3 Guidelines)



Enabling Role of Repository

The Repository is seen as the *enabler* of good Data Producer and Data Consumer practice

A data repository is designated a *Trusted Digital Repository* (TDR) if it:

- Enables Data Producers to adhere to Guidelines 1-3
- Meets guidelines 4-13
- Enables Data Consumers to adhere to Guidelines 14-16



Compliance

- Minimum level of compliance for each guideline
- Must be met to receive the Data Seal of Approval
- Compliance levels will be evaluated and will increase as:
 - Best practices emerge
 - Compliant tools become available
 - Implementation occurs



Compliance Levels

Level	Compliance Level Definition	Requirements
0	Not Applicable	Provide an explanation
1	We have not considered this yet	Provide an explanation
2	We have a theoretical concept	Provide a URL for the initiation document.
3	We are in the implementation phase.	Provide a URL for the definition document.
4	This guideline has been fully implemented for the needs of our repository	Provide a URL for the definition document.



Providing Evidence

- Important to link to publicly available documentation
- Documentation should be in English (or short summary in English)
- Topics for inclusion are suggested but they are neither exhaustive nor prescriptive
- Need adequate information for others to review the repository effectively



Peer Reviewer Guidance

- Does the self-assessment response correspond to the guideline?
- Are links to supporting documentation available publicly?
- Do you agree with the self-assessed compliance levels?
- Are they sufficient to award the DSA for this guideline?
- Have all abbreviations been explained?

In responding to the self-assessment, reviewers try to provide helpful comments and elicit additional information when needed.



DSA Governance

Community driven:

- **DSA Community:** all repositories holding a DSA
- **DSA General Assembly:** repositories holding a current DSA may choose to become member of the General Assembly. General Assembly members agree to perform a maximum of 3 peer reviews per year
- **DSA Board:** General Assembly members may put forward a person to be member of the DSA Board. Board membership is voted on by the General Assembly. The board governs the peer-review process, the modification/amendment of DSA guidelines and regulations, and general policy



Tour of the Guidelines



Guidelines – New for 2014-2015

- New “Guideline 0”: Repository Context. A brief general description of the functions and activities undertaken by the repository.
- Outsourcing now in principle possible for all guidelines, provided that the repository can prove sufficient level of control over the outsourced guideline
- Guideline 10 (The data repository enables the users to discover and use the data and refer to them in a persistent way) minimum level of compliance now 3 (was 2)



Guidelines for Data Producers: Content and Appraisal

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

- **Minimum:** We are in the implementation phase (3)
- Does the repository define the full package of information that should be deposited to facilitate assessment?
- What is included in a deposit?
- How much of the deposited information is passed on to the data consumers?



Data Producers: Data Formats

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

- **Minimum:** We are in the implementation phase (3)
 - Does the repository publish a list of preferred formats?
 - Ensure Data Producers adhere to the preferred formats?
 - Have a standard approach to deposits in non-preferred formats?



Data Producers: Documentation

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

- **Minimum:** Fully implemented (4)
 - Does the repository offer deposit forms or other user-friendly ways to submit metadata and what do they capture?
 - Use established metadata standards, registries, or conventions?
 - What is the repository's approach if the metadata provided are insufficient for long term preservation?



Data Repositories: Organizational Mandate

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

Minimum: Fully implemented (4)

- Does the repository have a Mission Statement? Under whose authority is the statement published?
- Describe how the Mission Statement is implemented?
- Have a succession plan in place for its digital assets?



Data Repositories: Compliance

5. The data repository uses due diligence to ensure **compliance with legal regulations and contracts.**

- **Minimum:** Fully implemented (4)
 - Does the repository use model contract(s)?
 - Publish conditions of use?
 - Have procedures for breach of conditions?
 - Ensure knowledge of and compliance with national and international laws? How?
 - Are special procedures applied to manage data with disclosure risk?



Data Repositories: Managing Storage

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

- **Minimum:** Fully implemented (4)
- Does the repository have a preservation policy?
- Have a strategy for backup / multiple copies? Please describe.
- Have data recovery provisions in place? What are they?
- How is deterioration of storage media handled and monitored?



Data Repositories: Preservation Plan

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

- **Minimum:** We are in the implementation phase (3)
 - 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.



Data Repositories: Workflows

8. Archiving takes place according to **explicit workflows** across the **data life cycle**.

➤ **Minimum:** We are in the implementation phase (3)

Does the repository have procedural documentation for archiving data? If so, provide references to:

- Workflows
- Decision-making for archival data transformations
- Data selection process
- Protecting privacy of subjects, etc.
- Approach toward data that do not fall within the mission



Data Repositories: Access

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

- **Minimum:** Fully implemented (4)
- Does the repository have licences / contractual agreements with data producers? Please describe.
- Enforce licences with the data producer? How?
- Have a crisis management plan? Please describe.



Data Repositories: Discovery, Use, and Identifiers

10. The data repository enables the users to **discover and use** the data and **refer to them** in a persistent way.

- **Minimum:** We are in the implementation phase (3)
- Does the repository offer search facilities?
- Are data provided in formats used by the Designated Community?
- Does the repository offer persistent identifiers?



Data Repositories: Integrity

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

- **Minimum:** We are in the implementation phase (3)
- Does the repository use checksums? What type?
- How is the integrity of data and metadata monitored?
- Does the repository deal with multiple versions of the data? If so, how? Please describe the versioning strategy.



Data Repositories: Authenticity

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

- **Minimum:** We are in the implementation phase (3)
 - Does the repository have a strategy for data changes? Are data producers made aware of this strategy?
 - Maintain provenance metadata and related audit trails?
 - Maintain links to metadata and to other datasets, and if so, how?
 - Compare the essential properties of different versions of the same file? How?



Data Repositories: Technical Infrastructure

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

- **Minimum:** We are in the implementation phase (3)
- What standards does the repository use for reference?
- How are the standards implemented? Please note any significant deviations from the standard with explanations.
- Does the repository have a plan for infrastructure development? Please describe.



Data Consumers: Access

14. The data consumer must comply with **access regulations** set by the data repository.

- **Minimum:** Fully implemented (4)
- Does the repository use End User Licence(s) with data consumers?
- 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?



Data Consumers: Codes of Conduct

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

- **Minimum:** Fully implemented (4)
- Does the repository need to deal with codes of conduct, including those designed for protection of research subjects?
- What are the terms of use to which data consumers agree?
- Are there measures in place to address breach of a code?
- Does the repository provide guidance in responsible use of confidential data?



Data Consumers

16. The data consumer respects the **applicable licences** of the data repository regarding the **use of the data**.

➤ **Minimum:** Fully implemented (4)

- Are there relevant licences in place?
- Are there measures in place to address licence breaches?



A Work in Progress

- These DSA Guidelines and their implementation are a work in progress and will evolve as further DSA assessments are performed
- We welcome your professional insight into this evolution either as a member of the DSA community or by directing your comments to info@datasealofapproval.org.



Procedures

Online tool: <http://www.datasealofapproval.org>

1. Applicant completes the initial application form and starts filling out the assessment. The Board assigns a reviewer.
2. Once the assessment is finished and submitted, the reviewer typically completes the review within two months.
3. If there are any issues, the reviewer sends the assessment back to the applicant with a request for clarification/ amendments and gives the applicant a deadline to respond.
4. If there are no further issues and all guidelines meet the minimum compliance level, the DSA is awarded!



How to Apply

Assessment | Data Seal of Approval

datasealofapproval.org/en/assessment/

Home **Assessment** Community News & Events TOOL LOG IN

Starting point for obtaining the Data Seal of Approval

Complete a self-assessment in the [DSA online tool](#). The online tool takes you through the 16 guidelines and provides you with support.

Once you have completed your self-assessment you can submit it for peer review. The peer reviewer will go over your answers and documentation and will award the Data Seal of Approval (DSA) when you have met the requirements.

Your self-assessment and review will not become public until the Data Seal of Approval is awarded.

Apply now!
or
Log in

Acquired Seals

Repository	DSA
3TU.Datacentrum	2010
Archaeology Data Service	2010
BABS - Long Term Preservation at the Bavarian State Library- Library Archiving and Access System	2010
Banco de Información para la Investigación Aplicada en Ciencias Sociales (BIIACS)	2014-2015
BAS CLARIN	2010
CLARIN-D Resource Center Leipzig	2010
CLARIND-UDS	2010
DANS: Electronic Archiving SYSTEM (EASY)	2010
Deutsches Textarchiv	2010
German National Library/ Deutsche Nationalbibliothek (DNB)	2010
HZSK Repository	2010
IDS Repository	2010
IMS Repository	2010

Receiving the Data Seal of Approval

After the Data Seal of Approval (DSA) is awarded by the Board, the DSA logo may be displayed on the repository's Web site. The Board will provide appropriate HTML code, which includes the DSA logo and a link to the organization's assessment



Next Steps

- Working with World Data System on harmonizing guidelines and procedures between the two organizations
- Research Data Alliance (RDA) working group established for this purpose
- Have created a first draft of common requirements for basic certification
- Ultimately may lead to wider global framework for certification
- See rd-alliance.org for working group outcomes



Questions?