



Data Seal of Approval

Paul Trilsbeek

The Language Archive, MPI for Psycholinguistics

CLARIN Center Assessment Workshop Copenhagen, 25 April 2013



Outline

- DSA History, context, objectives, organization, ...
- Guidelines walkthrough



History

- Initiated by DANS (Den Haag, Netherlands) as national “datakeurmerk” in 2005, first version presented in 2007
- Internationalised and handed over to international board in 2009
- Now part of a proposed European Framework for Audit and Certification



European Framework for Audit and Certification

3. Formal certification: DSA + full external audit and certification based on ISO 16363* or DIN 31644**
2. Extended certification: DSA + structured, externally reviewed and publicly available self-audit based on ISO 16363* or DIN 31644**
1. Basic certification: Data Seal of Approval (DSA)

<http://www.trusteddigitalrepository.eu>

*ISO 16363 - Audit and Certification of Trustworthy Digital Repositories

**DIN 31644 - Information and Documentation - Criteria for Trustworthy Digital Archives



DSA Overview

- 16 Guidelines for Trusted Digital Repositories
- Guidelines that relate to Data Producers (3), Data Repository (10) and Data Consumer (3)
- Some guidelines can be outsourced, others not
- Self-assessment, no external auditors or site visit
- Peer-reviewed process supervised by DSA Board
- DSA granted for a period of max. 2 years
- Online tool for self-assessment and review



Current DSA Board

- Henk Harmsen (Chair. DANS, Netherlands)
- Hervé L'Hours (UK Data Archive, United Kingdom)
- TBD (CINES, France)
- Sabine Schrimpf (NESTOR/DNB, Germany)
- Paul Trilsbeek (TLA/MPI, Netherlands)
- Mary Vardigan (ICPSR, United States)



Foundations

- Trust
- Transparency
- Community



Objectives

Data Producers

- *Assurance* of reliable data Storage

Funding Bodies

- *Confidence* that data is available for re-use

Data Consumers

- *Enables* assessment of repositories



Principles

The data are:

- available on the Internet
- accessible
 - while taking into account relevant legislation with regard to personal information and intellectual property of the data.
- usable (file formats)
- reliable
- citable (can be referred to)



Stakeholders

Data Producer

- responsible for the quality of the digital data

Data Repository

- responsible for the quality of data storage & availability

Data Consumer

- responsible for the quality of use of the digital data



Responsibility

But... the DSA focus is on the Repository as *enabler* of good Data Producer and Data Consumer practice

A data repository is designated a *Trusted Digital Repository* (TDR) according to the requirements of the Data Seal of Approval if it *meets* guidelines 4-13 and if it *enables* data-producers and data-consumers to adhere to guidelines 1-3 and 14-16.

The application for the Seal must include a web address, clearly associated with the repository, where the Seal will be displayed.



Compliance

For each DSA Guideline there is a minimum level of compliance reflecting how mature practices should be for a repository to be awarded the Seal. As best practices emerge, compliant tools become available and implementation occurs the DSA Board will re-evaluate the minimum compliance level.

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.



Evidence

Transparency

- Link to publicly available documentation
 - Or deadline for public release
- English or short summary in English

Reviewers Guide: “Topics for discussion and inclusion are suggested but they are neither exhaustive nor prescriptive”

How do we know what is

- Appropriate?
- Sufficient?



Peer Reviewers

Guidance asks them to consider:

- Does the supplied information correspond to the guideline in question? (Also refer to the applicant supporting text for each guideline)
- Are the links to supporting documentation available publically?
- Do you agree with the self-assessed compliance levels and are they sufficient to award the DSA for this guideline?
- Have abbreviations been explained
- In responding to the self-assessment try to provide helpful comments rather than specific questions.



Data Producers

Guidelines 1 to 3 relate to the level of guidance which the repository gives to the data producer before and at the time of submission to the repository.

The responses should concentrate on the contribution of the repository in making compliance with this guideline possible for the data producer.



Data Producers: the Content

“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)
- **Outsourced:** No



Data Producers: the Content

“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.”

- Sector-specific/Designated Community quality criteria
- Transparency
- Adherence to disciplinary & ethical norms
- Assessment by experts and colleagues



Data Producers: the Content

“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.”

- 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 official approval for data collection to confirm adherence to legal or ethical requirements?
- Does the repository promote data sharing and reuse?
- Does the repository provide enough information to answer questions such as the following?
 - Are the data based on work performed by the data producer (researcher or institution that makes the available) and does the data producer have a solid reputation?
 - Data regarding, or references to, the data producer and/or their affiliated organisation(s)
 - References to publications pertaining to the submitted data
 - Was data collection or creation carried out in accordance with legal and ethical criteria prevailing in the data producers geographical location or discipline (e.g. Ethical Review Committee or Data Protection legislation)?
 - Information regarding the methods and techniques used, including those for data collection.
 - Are the data suitable for reuse?
 - Information regarding the manner in which the research data have obtained shape , content and structure



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)
- **Outsourced:** No

Obsolete formats create a risk of unusable data

Preferred formats that a data repository can reasonably assure will remain readable and usable

Usually De-Facto Standards



Data Producers: Data Formats

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

- 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?



Data Producers: Documentation

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

- **Minimum:** Fully implemented (4)
- **Outsourced:** No

Context through metadata

be provided in accordance with the guidelines of the data repository

- *Descriptive metadata*
- *Structural metadata*
- *Administrative metadata data.*

The data repository specifies the level of producer-created metadata required and provides the tools for its effective capture.



Data Producers: Documentation

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

- 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 file level?
- 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 preservation?



Data Repositories: Organisation and processes

Organizations that play a role in digital archiving and are establishing a *Trusted Digital Repository* minimally possess a sound, long-term financial, organisational and legal basis



Data Repositories: Organisation and processes

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

- **Minimum: Fully implemented (4)**
- **Outsourced: Yes**



Data Repositories: Organisation and processes

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

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 related 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?



Data Repositories: Organisation and processes

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

- **Minimum:** Fully implemented (4)
- **Outsourced:** No



Data Repositories: Organisation and processes

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

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

- Is the repository a legal entity? 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 anonymise files or to provide access in a secure way?
- Are staff trained in management of data with disclosure risk?



Data Repositories: Organisation and processes

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

- **Minimum: Fully implemented (4)**
- **Outsourced: Yes**



Data Repositories: Organisation and processes

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

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?



Data Repositories: Organisation and processes

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

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



Data Repositories: Organisation and processes

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

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?



Data Repositories: Organisation and processes

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

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



Data Repositories: Organisation and processes

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

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?



Data Repositories: Organisation and processes

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

- **Minimum: Fully implemented (4)**
- **Outsourced: No**



Data Repositories: Organisation and processes

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

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.



Data Repositories: Organisation and processes

“10. The data repository enables the users to utilize the data and refer to them.”

- **Minimum:** We have a theoretical concept (2)
- **Outsourced:** No



Data Repositories: Organisation and processes

“10. The data repository enables the users to utilize the data and refer to them.”

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

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



Data Repositories: Organisation and processes

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

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** No

The information contained in the digital objects and metadata is complete and does not change relative to the originals.



Data Repositories: Organisation and processes

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

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.



Data Repositories: Organisation and processes

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

- **Minimum:** We are in the implementation phase (3)
- **Outsourced:** No

This pertains to the degree of reliability of the original and to the provenance of the data. Existing relationships between data sets and explicit links are maintained.



Data Repositories: Organisation and processes

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

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?



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)
- **Outsourced:** Yes

The technical infrastructure constitutes the foundation of a Trusted Digital Repository. The OAIS reference model, an ISO standard, is the de facto standard for using digital archiving terminology and defining the functions that a data repository fulfils



Data Repositories: Technical Infrastructure

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

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

- 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.

Or if applicable

- To which TDR have you outsourced?



Data Consumers

The data consumer uses the digital research data in compliance with guidelines 14-16

The quality of the use of research data is determined by the degree to which the data can be used without limitation for scientific and scholarly research by the various target groups, while complying with certain applicable codes of conduct.

The open and free use of research data takes place within the relevant legal frameworks and the policy guidelines as determined by the relevant national authorities.

With regard to accessing information, the data consumer is bound by relevant national legislation. The data repository may have separate access regulations, which include restrictions imposed by the laws of the country in which the data repository is located. Access regulations should be based on relevant international access standards (e.g., Creative Commons) as much as possible.

Most nations have legal frameworks relating to the ethical use and re-use of data. These frameworks range from the statutory — which protect the privacy of individuals — to formal codes of conduct which inform ethical issues. Repositories must be aware of these local legal frameworks and ensure that they are taken into account when providing data for re-use.



Data Consumers

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

- **Minimum: Fully implemented (4)**
- **Outsourced: No**



Data Consumers

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

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?



Data Consumers

“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)**
- **Outsourced: No**



Data Consumers

“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.”

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?



Data Consumers

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

- **Minimum: Fully implemented (4)**
- **Outsourced: No**



Data Consumers

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

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?



DSA for CLARIN Centers

“Centres need to have a proper and clearly specified repository system and participate in a quality assessment procedure as proposed by the Data Seal of Approval or MOIMS-RAC approaches”

CLARIN Center requirements for repositories have quite a bit of overlap with DSA guidelines

CLARIN Center repositories that fulfill CLARIN requirements should be able to apply for DSA with relatively little effort, provided that sufficient documentation is in place

DSA is mainly about being explicit and transparent about procedures, implementations and goals



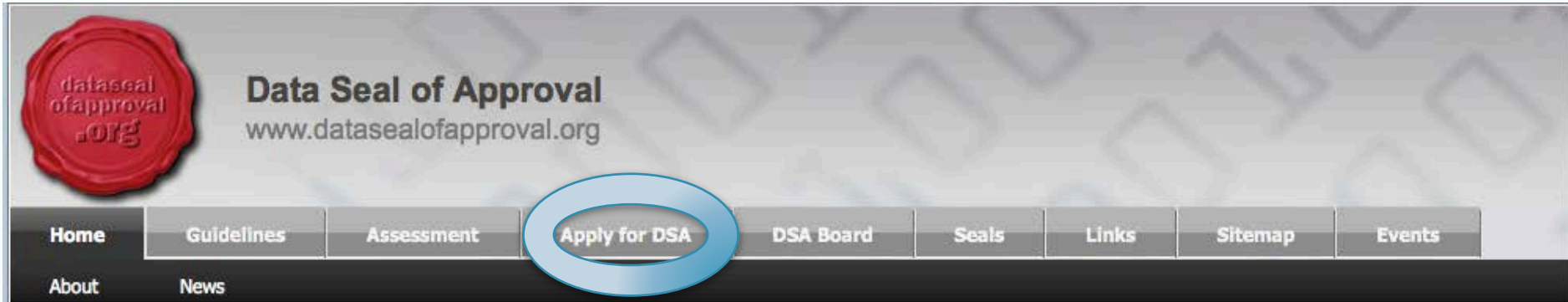
DSA Guidelines

Online tool:

- <http://www.datasealofapproval.org>



How to apply



The screenshot shows the website's navigation menu. The "Apply for DSA" link is highlighted with a blue oval. The menu items are: Home, Guidelines, Assessment, Apply for DSA, DSA Board, Seals, Links, Sitemap, Events, About, and News.

Data Seal of Approval
www.datasealofapproval.org

Home Guidelines Assessment **Apply for DSA** DSA Board Seals Links Sitemap Events
About News

Info

About the Data Seal of Approval (DSA)

The Data Seal of Approval ensures that in the future, research data can still be processed in a high-quality and reliable manner, without this entailing new thresholds, regulations or high costs. The Data Seal of Approval and its quality guidelines may be of interest to research institutions, organizations that archive data and to users of that data. It can be granted to any repository that applies for it via the [assessment procedure](#).

Anyone who archives his or her data would like to be able to find, recognise and use it in the future. With electronic data this cannot be taken for granted, after all hardware and software are changing all the time. Making data future-proof can be accomplished by ensuring that data sets and metadata meet certain requirements. In consultation with large data producers and managers, the Dutch data archive DANS documented what those requirements need to be in the Data Seal of Approval, which have been further developed since and handed over to the DSA Board in May 2009.

You can download the document that contains the quality guidelines of the Data Seal of Approval below.

Contact: info@datasealofapproval.org



A Work in Progress

Please note that the Guidelines and their implementation are a work in progress which 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.