Data Seal of Approval

dissemination, assessment & procedures in 30 slides

Henk Harmsen
Deputy director DANS

henk.harmsen@dans.knaw.nl

www.dans.knaw.nl
Data Seal of Approval - INDEX

– DANS task
– Data Seal of Approval history
– Roles and responsibilities of stakeholders
– Data Seal of Approval in general
– The 16 guidelines
– Control or trust?
– Assessment?
– The DSAA-IB
– The roadmap
Data Archiving & Networked Services

– an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), and is also supported by the Netherlands Organization for Scientific Research (NWO).

– since its establishment in 2005, DANS has been taking care of storage and continuous accessibility of research data in the social sciences and humanities.
**TASK**

- Blueprint DANS 2004:
- Make Dutch data-archives ‘DANS proof’. Archives will be judged on the base of ‘DANS standards’ and eventually they will be upgraded according these norms.
- If necessary or desirable DANS will take initiatives to make existing data-infrastructure ‘DANS proof’
Quality criteria research data

Research data

1. can be found on the internet
2. are accessible
3. are in a usable format available
4. are trustworthy
5. can be referred to
Reinventing the wheel?

- **Kriterienkatalog vertrauenswürdige digitale Langzeitarchive** - NESSTOR
- **Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist** - Research Library Group (CRL) - 84 criteria
- **Digital Repository Audit Method Based on Risk Assessment** (DRAMBORA) - Digital Curation Centre (DCC) and Digital Preservation Europe (DPE)
- Digital Preservation Management (DPM) framework 2.0 ICPSR ...
Back to the basics

DANS standards -> Trusted Digital Repositories
DANS-proof -> Data Seal of Approval
repository -> stakeholders
certification -> guidelines
audit -> self-assessment
control -> trust
details -> overview
Stakeholders

– funding agencies
– data producers
– data consumers
– data repositories
Roles

- Funding agencies: oblige seal of approval
- Data producers: produce excellent data
- Data consumers: use data in a trustworthy way
- Data repositories: responsible for long term preservation and distribution of research data
Responsibilities

– The *data producer* is responsible for the quality of the digital research data.

– The *data repository* is responsible for the quality of the storage and the availability of the data: data management.

– The *data consumer* is responsible for the quality of use of digital research data.
The data producer (1-3)

Guidelines relating to the quality of digital research data

1. The data producer **deposits** the research **data** in a data repository **qualified** according to these guidelines.
2. The data producer provides the research **data** in **formats recommended** by the data repository.
3. The data producer provides the research **data together with the metadata requested** by the data depository.
The data repository (4-13)

Guidelines relating to the quality of the data repository

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

5. A data repository **ensures** that **legal regulations and contracts** are complied with.

6. A data repository applies processes and procedures for **guaranteeing quality management** for the storage of data.

7. The data repository has a **long-term plan** regarding measures for **durable archiving**.

8. Archiving takes place according to **predetermined criteria**.

9. A data repository **assumes responsibility** from the data producers for **access** and **availability** of the digital objects.
10. A data repository enables the users to use the research data and refer to them.

11. A data repository ensures the integrity of the digital objects and the metadata. The information contained in the digital objects and metadata is complete and does not change relative to the originals.

12. A data repository ensures the authenticity of the digital objects and the metadata. This regards the degree of reliability of the originality and origin. Existing relationships between data sets and explicit links are maintained.

13. The technical infrastructure explicitly supports the tasks and functions as mentioned in the OAIS reference model.

14. The technical infrastructure of a data repository supports guidelines 4 through 12.
The data consumer (14-16)

*Guidelines relating to the quality of the use of digital research data*

14. With regard to accessing information, the data consumer is bound by national legislation. The data repository has a separate access regulation, which includes the restriction 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. The data consumer must comply with the applicable frameworks and access regulation.
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 of knowledge and information.

16. The data consumer respects the applicable licenses of the data repository with regard to use of the research data.
Indeed a repository, but not an archive, still in for the data seal of approval?

– Implement eleven guidelines: 1, 2, 3, 5, 9, 10, 11, 12, 14, 15 en 16.

– Let five guidelines (4, 6, 7, 8 and 13) be executed by another DSA archive.
Implement: 1-3, 5, 9-12 and 14-16

- make use of a qualified data repository
- ingest research data formats as recommended by the data repository
- ingest metadata as recommended by the data repository
- ensure legal regulations and contracts
- ensure access and availability
- ensure the use of and references to data
- ensure the integrity of the digital objects and the metadata
- ensure the authenticity of the digital objects and the metadata.

- respect national legislation
- respect codes of conduct in higher education and scientific research
- respect the applicable licenses of the data repository
Subcontract: 4, 6, 7, 8 and 13

- explicit mission
- guarantee for quality management
- long-term plan for durable archiving.
- archiving according to predetermined criteria
- technical infrastructure international standards compatible
Unique selling points

- DSA is oriented toward **scientific data**, not primarily toward publications.
- DSA not only pays attention to the archiving institution, but also to the data producer and the data consumer (**shared responsibility**).
- DSA is not in conflict with for example TRAC, but is rather a step **toward** it.
- DSA does not chooses standardization but opts for ‘**trust**’ (like custom of peer review in the scientific world).
- DSA also focuses on **smaller organizations**.
- DSA is relatively **light** and therefore easy to implement.
- **Openness, dynamics** and **speed** are possible in the actual implementation.
- DSA is formulated as **points of attention, not** as **solutions**.
- DSA offers possibilities for **subcontracting archiving** and still meet the requirements of the DSA.
Trusted Digital Repository

According to the DANS blueprint

- any digital data archive that is ‘DANS-proof’

Now we say:

- Any digital data archive with an infrastructure which has the 16 guidelines of the Data Seal of Approval successfully implemented; shown by a public self assessment and an interpretation of these guidelines.
Control?

- Certification
  - based on national (NEN) or international (ISO) standards: takes long and hard to change
  - certifying agency
  - audits
  - = costs a lot!!
  - ... and what does it say?
Trust?

– yes
  – national
  – International

– with some minor instruments for the stakeholders:
  – *Quality of research data*: peer reviews via the internet
  – *Quality of the data repository*: self assessment (DSA, DRAMBORA, TRAC or combination) and low level audit (trustworthy (international) colleague archives or agency)
  – *Quality of use of data*: in case of abuse (rarely), disciplinary action by employer (universities / funding agencies).
Quality of the data repository

– First step: Awareness. **Data Seal of Approval Assessment** (DSAA).
– Building blocks: keystone...
– After the awareness: DRAMBORA, TRAC
Quality of the data repository

What is the **DSAA**?

- A description, which can be found on your repository page, of how you implemented the 16 DSA guidelines
- An assignment of the DSA logo by the DSAA international board (DSAA-IB)
DSAA – International Board

(Henk Harmsen), Paul Trilsbeek (NL, MPI/CLARIN), Natascha Schumann (D, DNB/NESTOR), Olivier Rouchon (F, CINES), Mary Vardigan (USA, ICPSR), (Lisa de Leeuw), Hans Pfeiffenberger (D, AWI, IPYDLE), Matthew Woolard (UK, UKDA/CESSDA), Laurents Sesink (NL, DANS/DARIAH).
The board as an editorial board

- Search, assess and teach new reviewers
- Bright expertise on different domains
- Responsible for the DSA guidelines
- Responsible for the DSAA guidelines and implementation
Step-by-step

- Self-assessment of the repositories of the ISAA-IB members according to the 16 guidelines
- On the agenda of international projects:
  - Planets (via the National Archives en the Royal Library), DRIVER II (via SURF), DARIAH, CLARIN, CESSDA... (via the DSAA IB)
  - Funding agencies (NWO, SURF, KNAW)
  - Digital coalitions like DCC and NCDD
  - Dialogue with Nestor, Drambora and Trac... (this workshop)
- European workshop May 2009 (hosted by the EC), domain map ready, handover DSA responsibility to the DSAA-IB (may 2009)
- Assessments of other repositories
Data Seal of Approval:

In 29 slides...

Thank you!

www.datasealofapproval.org
www.datakeurmerk.nl/en