Trust and Certification: the case for Trustworthy Digital Repositories

RDA Europe webinar, 14 February 2017
Ingrid Dillo, DANS, The Netherlands
“Perhaps the biggest challenge in sharing data is trust: how do you create a system robust enough for scientists to trust that, if they share, their data won’t be lost, garbled, stolen or misused?”

The Data Harvest: How sharing research data can yield knowledge, jobs and growth

An RDA Europe Report
December 2014
Trust is at the very heart of storing and sharing data

- Users
- Depositors
- Funders
What is trust built on?

- Dedicate yourself (mission statement)
- Do what you promise (stable, sincere and competent reputation)
- Be transparent (peer review, get certified)

"They don’t trust each other to share research."
Any organisation which provides access to data over a long period of time should be fully trusted only with a public statement describing the practices they follow and the provenance of data they provide. Standards of trust are critical.
Building blocks of trustworthiness

- actions and attributes of the trustee (integrity, transparency, competence, predictability, guarantees, positive intentions)

- external acknowledgements:
  - reputation (researchers),
  - third party endorsements (funders)
Certification of digital repositories

Standards can play an important role in establishing trust
European framework levels

- **Core Certification** is granted to repositories which obtain DSA certification

- **Extended Certification** is granted to repositories which perform a structured, externally reviewed and publicly available self-audit based on DIN 31644

- **Formal Certification** is granted to repositories which obtain full external audit and certification based on ISO16363
Global framework

Formal

Extended

Core
DSA key characteristics

- Basic light-weight certification standard
- 16 Guidelines for Trustworthy Digital Repositories
- Self-assessment, no external auditors or site visit
- Peer review process supervised by international DSA Board
- Online tool for self-assessment and review
- DSA granted for a period of 2 years
- Open, transparent and inclusive (public self assessment)
- Focus on social sciences and humanities
- Strong in Europe (CESSDA, CLARIN, DARIAH, EUDAT)
- Around 65 seals acquired
WDS key characteristics

- World Data System part of ICSU
- Light-weight certification procedure for regular and network members
- Based on self assessment
- Peer review by WDS Scientific Committee
- Focus on earth and spatial sciences
- Many members in US and Asia
- Renewal between 3 and 5 years
- Some 70 accredited members
DIN 31644: extended certification

- 34 criteria written by German NESTOR-group and adopted in Germany as DIN31644
- Self-assessment procedure by NESTOR leads to NESTOR seal
- Review of the assessment by 2 reviewers, appointed by NESTOR
- Self assessment and evidence on website
- 2 seals acquired (DANS and DNB)

http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.htm
ISO 16363: formal certification

- Based on Open Archival Information System (OAIS) and Trusted Repository Audit and Certification (TRAC)
- Over 100 metrics
- Test audits 2011 by PTAB (Primary Trustworthy Digital Repository Authorisation Body)
- Full external auditing process
- ISO 16919: Requirements for bodies providing audit and certification of candidate trustworthy digital repositories
- No formal ISO certifications yet..

http://www.iso16363.org/
DSA and WDS: look-a-likes

- **Communalities:**
  - Lightweight, community review

- **Complementarity:**
  - Geographical spread
  - Disciplinary spread
• Research Data Alliance: aims to build the social and technical bridges that enable open sharing of data

• WGs and IGs working on a large variety of topics

• WG consisting of representatives from both organizations has explored and developed a DSA–WDS Partnership (18 months)
Partnership goals

- Realizing efficiencies
- Simplifying assessment options
- Stimulating more certifications
- Increasing impact on the community
Working Group outcomes

- Common catalogue of requirements for core repository assessment
- Common procedures for assessment
- Shared testbed for assessment
New common requirements

- Context (1)
- Organizational infrastructure (6)
- Digital object management (8)
- Technology (2)
- Additional information and applicant feedback (2)
Requirement compliance levels

- 0 – Not applicable
  - 1 – The repository has not considered this yet
  - 2 – The repository has a theoretical concept
  - 3 – The repository is in the implementation phase
  - 4 – The guideline has been fully implemented in the repository

- .. to foster the applicants’ own understanding of the current status/maturity of their repositories
Context

• Repository type

• Brief description of the repository’s designated community

• Level of curation performed

• Outsource partners
R1. The repository has **an explicit mission** to provide access to and preserve data in its domain.

R2. The repository maintains all applicable **licenses** covering data access and use and monitors compliance.

R3. The repository has a **continuity plan** to ensure ongoing access to and preservation of its holdings.

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with **disciplinary and ethical norms**.

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.

R6. The repository adopts mechanism(s) to secure ongoing **expert guidance** and feedback (either in-house, or external, including scientific guidance, if relevant).
R7. The repository guarantees the integrity and authenticity of the data.

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

R9. The repository applies documented processes and procedures in managing archival storage of the data.

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.
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.

R12. Archiving takes place according to defined workflows from ingest to dissemination.

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

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.
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.

- R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.
Digital Object Management

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

**Compliance Level:**

<table>
<thead>
<tr>
<th>Response</th>
</tr>
</thead>
</table>

**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., integrity 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 following 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.
New requirements are out now!

WDS and DSA Announce Unified Requirements for Core Certification of Trustworthy Data Repositories developed through the RDA DSA–WDS partnership Working Group


Parallel Assessment Processes:

- URLs to evidence strongly encouraged
- Maturity ratings strongly encouraged
- Assessments to be publicly available
- Successful completion means certification in both DSA and WDS
- Renewals every three years
Common procedures

- Sustainable Review Process
  - Pool of reviewers (training provided) drawn from DSA and WDS
  - Two reviewers (from DSA and WDS) for each application, approved by the new DSA–WDS Certification Board

- Mutual Governance Process
  - DSA and WDS agree to work together to implement and steward the partnership
Testbed

- 2 DSA applicants
- 4 WDS applicants
- Overall positive results
- Minor tweaks
- More explanations
Why do we do this at DANS?

- Certification as a means to build trust in our repository with our clients, both depositors and users of data, with our partner organizations and with research funders.

- Certification as a big stick to further develop and professionalize our core services, workflows and our organization as a whole.
Benefits and value of core certification (as noted by DSA-repositories)

- Builds stakeholder confidence in the repository
- Improves communication within the repository
- Improves repository processes
- Ensures transparency
- Differentiates the repository from others
- Saves time and labor over other certification methods
Thank you very much for listening!

ingrid.dillo@dans.knaw.nl

www.dans.knaw.nl
http://datasealofapproval.org/en/
https://www.icsu-wds.org/services/certification