Trust and Certification: An Introduction to the DSA–WDS Common Requirements

Ingrid Dillo
Deputy Director DANS / Vice Chair WDS-SC
ingrid.dillo@dans.knaw.nl

WDS Webinar #10, 31 August 2016
Webinar overview

Goal

Introduction to new Common Requirements for the certification of WDS Regular Members

First possibility to prepare for WDS Members’ Forum discussion on the Requirements in Denver (11 September)
Webinar overview

Topics

• The importance of trust and certification
• Certification levels and the global framework
• Core level certification and the new Common Requirements
“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?”
Trust

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)
What is a trustworthy digital repository?

Things are not always what they say they are. Things do not always state what they are.
Trust in repositories: An example

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.
Certification of digital repositories

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

**Core Certification** is granted to repositories that obtain DSA certification.

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

**Formal Certification** is granted to repositories that obtain full external audit and certification based on ISO16363.
Global framework
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/
WDS: Core certification

- World Data System part of ICSU
- Light-weight certification procedure for regular and network members
- Based on self-assessment
- Peer review overseen by WDS Scientific Committee
- Focus on earth and spatial sciences
- Many Members in US and Asia, as well as a number in Europe
- Renewal between 3 and 5 years
- Over 70 accredited members
DSA: Core certification

- Basic light-weight certification standard
- 16 guidelines for Trustworthy Digital Repositories
- Guidelines that relate to Data Producers (3), Data Repository (10) and Data Consumer (3)
- 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)
- Over 60 seals acquired, some 50 in progress
DSA and WDS: Look-a-likes

Communalities:
• Lightweight, community review

Complementarity:
• Geographical spread
• Disciplinary spread
The umbrella of the RDA

- 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
WG activities

• Develop Common Catalogue of Requirements for Core repository assessment
• Develop Common Procedures for assessment
• Implement a shared testbed for assessment
• Ultimately, create a shared framework for certification that includes other standards as well, including DIN/nestorSeal and ISO 16363
New Common Requirements: Outline

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

https://rd-alliance.org/system/files/DSA–WDS%20Catalogue%20of%20Common%20Requirements%20V2.2.pdf
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. If applicable, please list them
Organizational infrastructure

I. Mission/Scope

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

II. Licenses

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

III. Continuity of access

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

IV. Confidentiality/Ethics

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

V. Organizational infrastructure

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.

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant)
Digital Object Management

VII. Data integrity and authenticity

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

VIII. Appraisal

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

IX. Documented storage procedures

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

X. Preservation Plan

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

XI. Data Quality

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.

XII. Workflows

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

XIII. Data discovery and identification

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

XIV. Data reuse

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.
Technology

XV. 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.

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.
Common Requirements
Organizational Infrastructure
IV. Confidentiality/Ethics

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

Compliance Level

Response

Guidance:
Adherence to ethical norms is critical to responsible science. Disclosure risk—for example, the risk that an individual who participated in a survey can be identified or that the precise location of an endangered species can be pinpointed—is a concern that many repositories must address.

For this Requirement, responses should include evidence related to the following questions:

- How does the repository comply with applicable disciplinary norms?
- Does the repository ensure that data collection or creation was carried out in accordance with legal and ethical criteria prevailing in the data producer’s geographical location or discipline (e.g., Ethical Review Committee/Institutional Review Board or Data Protection legislation)?
- 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 the management of data with disclosure risk?
- Are there measures in place if conditions are not complied with?
- Does the repository provide guidance in the responsible use of disclosive, or potentially disclosive data?

Evidence for this Requirement should be in alignment with provisions for the procedures stated in R12 (Workflows) and for any licenses in R2 (Licences).
Comparison with current WDS requirements

• No huge differences
• Slightly changed structure
• Some different accents
• More emphasis on documented procedures and plans
Common procedures

Parallel Assessment Processes:

- URLs to evidence strongly encouraged
- Maturity ratings strongly encouraged
- Assessments to be publicly available
- 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
In the coming months…

• Both DSA and WDS will introduce the new Common Requirements: Launch at IDW

• Discussion on the Requirements at the WDS Members’ Forum in Denver

• Later this year, a follow-up Webinar looking at each of the Requirements in more detail: What is expected as an applicant and a reviewer?

• WDS will (hopefully) start working with the new Requirements on 1st October

• One common standard body; two different certification authorities
Further into the future…

• Continue work on the common procedures, with the ultimate goal to create a joint authority to bestow the DSA and WDS certification at once

• Conduct outreach to other standards like nestorSeal and ISO 16363
Thank you for listening

ingrid.dillo@dans.knaw.nl

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