Core Trustworthy Data Repository Requirements

Ingrid Dillo, Deputy Director DANS

Webinar CENDI, 13 February 2017
Two aspects highlighted

• What was the **philosophy** or the **process** whereby you came up with the Core across the various schemes?

• What are the **experiences** of yourselves and others with regard to implementing the criteria, particularly if you put this in the context of actual certification? What are the best practices, issues, lessons learned, etc.?
The certification landscape
DSA: core certification

- 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 Board
- DSA granted for a period of 2 years
- Focus on social sciences and humanities
- Strong in Europe (CESSDA, CLARIN, DARIAH, EUDAT)
- around 70 seals acquired
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 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
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
1. Organisational infrastructure

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).
2. Digital object management (1)

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.
2. Digital object management (2)

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.
3. 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:

Response

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., ftxity 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!


Experiences at DANS

- Data service provider (50 colleagues); institute of KNAW and NWO; 2005 and 1964
- Mission to promote permanent access to digital research data

2011: certification of the DANS repository becomes an important target in long term policy strategy
2011: DSA seal
2011: ISO test audit
2013: renewal DSA seal
2014-2015: nestorSeal, WDS accreditation
2017: new core requirements
Certification in practice: broad scope

- Broad range of topics: organizational, staffing, financial and legal aspects, archival workflows, IT-infrastructure, risk management, etc.

- Properly describing policies, processes, etc.

- Development of missing policies, processes, IT- and infrastructural elements, etc.
Certification in practice: organizational aspects

• Responsibility for achieving the target on management level

• Core certification team: planning, discussing, monitoring and partly executing the work

• Many colleagues within DANS with specific expertise temporarily involved in the actual work
Certification in practice: effort involved

- Highly dependent on your level of entry
- The effort will rise, if you still need to do real work in order to comply with the guidelines
- The effort will rise when you climb the certification stairs to an extended and formal certification level
Effort DSA-renewal

In total around 250 hours of work:

- policies: 26 hours
- technical development: 106 hours
- writing the self-assessment: 98 hours
- project management: 16 hours
NCDD survey 2016: level of investment

Orientation phase:  
80% of respondents estimated an investment of up to 20 hours.

Preparation (self-assessment) phase:  
50% up to 100 hrs.; 30% up to 200 hrs.

Peer review process:  
50% up to 50 hrs.; 30% up to 100; 12% up to 200 hrs.
Why do repositories invest in certification efforts?

• Builds stakeholder confidence in the repository (funders, publishers, etc.)
• Raises awareness about digital preservation
• Improves communication within the repository
• Improves repository processes
• Ensures transparency
• Differentiates the repository from others
NCDD survey 2016: perceived benefits

“The benefits propagated by the DSA itself are in line with the perception of the respondents: this is most clearly the case with the stated benefit “awareness raising about digital preservation,” followed by “stakeholder confidence.”

“When queried about other perceived benefits, it is clear that the certification process not only led to external benefits but also to improved internal processes, documentation and opportunities to attract data producers as well as data consumers.”

Quote: “The experience of applying and the issues that came up during the process have turned out to be very positive and are helping us consolidate our quality related working lines.”
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
“The respondents perceive the benefits of DSA-certification as both tangible and critical to the continuing fulfillment of their mission.”

“The majority of the respondents rated the ratio between investments and benefits as “adequate-rewarding” to “rewarding-excellent.”

“Almost no one aimed for certification at a higher level (DIN 31644, ISO 16363).”
Main takeaways

• Start with an internal quick scan to determine the level of entry against the catalogue of requirements

• Commitment from the top is crucial

• Broad support within the organization is needed

• Use the framework: do not aim to high at once
Thank you for listening

ingrid.dillo@dans.knaw.nl
