CoreTrustSeal for Trustworthy Data Repositories

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

WDS Data Repositories Day, Gaborone, 9 November 2018
Topics

• Why certification is important
• The different certification standards
• CoreTrustSeal: requirements and procedures
• Certification in practice
• The benefits of certification
INTERNATIONAL ACCORD ON OPEN DATA FOR OPEN SCIENCE

Preface
The International Council for Science (ICSU), the Inter-Academy Partnership (IAP), The World Academy of Sciences (TWAS) and the International Social Science Council (ISSC) have created a joint enterprise, Science International, to be the global science community’s voice of policy for science. This accord is its first foray in that domain. The accord identifies the challenges and opportunities of the global data revolution as the predominant issue of current policy for science. It wishes to add the distinctive voice of the scientific community to those of governments and inter-governmental bodies that support science and research, and of data as a fundamental pre-requisite if science is to remain a public good.

Realising the European Open Science Cloud

First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud

M A K I N G O P E N S C I E N C E A R E A L I T Y

A F R I C A O P E N S C I E N C E H A R D W A R E
THE DATA ARE MINE!

IVORY TOWER OF SCIENCE

SCENE FROM THE PAST?
“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?”
Data repositories

“When asked where they have published data, most commonly respondents had done so as an appendix to an article (just over 30%) with a data repository close behind (just under 30%) and 20% having published in a data journal.

36% of respondents have lost data on which they were working and there is, unsurprisingly, a high correlation between the vehicle for storing data and where it was lost - computer hard drives were the most common culprit here.”

The pillars of trust

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

• external acknowledgements:
  – reputation (researchers)
  – third party endorsements (funders, publishers)
Certification of data repositories

Certification can play an important role in establishing trust

Framework of certifications at different levels
The global certification landscape

Formal
http://www.iso16363.org/

Extended
DIN 31644
http://www.dnb.de/Subsites/nestor/EN/Siegel/siegel.html

Core
https://www.coretrustseal.org
Formal certification: ISO 16363

- 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, 5 year renewal cycle
- ISO 16919: Requirements for bodies providing audit and certification of candidate trustworthy digital repositories

http://www.iso16363.org/
PTAB has awarded the World’s first ISO 16363 certification to the Indira Gandhi National Centre for the Arts

PTAB was the first in the World to be accredited to perform ISO 16363 Audit and Certification

PTAB, incorporated by the same experienced international group of digital preservation experts who developed ISO standards 14721, 16363 and 16919, has been accredited by the National Accreditation Board for Certification Bodies of India (NABCB) to conduct ISO 16363 audits worldwide utilizing ISO standard 17021, as extended by ISO 16919. ISO processes ensure that any approved audit organisation accredited by a National Accreditation Board, may conduct audits worldwide and any certification that body grants, is accepted worldwide.
Extended Certification: nestorSeal

- 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
- No renewal cycle

http://www.dnb.de/Subsites/nestor/EN/Siegel/siegel.html
## EVALUATED ARCHIVES

<table>
<thead>
<tr>
<th>Archive</th>
<th>Year</th>
<th>Link to documentation</th>
<th>Application forms</th>
<th>Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIB Technische Informationsbibliothek Hanover</td>
<td>2017</td>
<td>Documentation</td>
<td>Application form</td>
<td>Report</td>
</tr>
<tr>
<td>Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZBW Leibniz-Informationszentrum Wirtschaft</td>
<td>2017</td>
<td>Documentation</td>
<td>Application form</td>
<td>Report</td>
</tr>
<tr>
<td>Data Archiving and Networked Services (DANS)</td>
<td>2016</td>
<td>Documentation</td>
<td>Application form</td>
<td>Report</td>
</tr>
<tr>
<td>Deutsche Nationalbibliothek - für den Bereich Netzpublikationen -</td>
<td></td>
<td>Documentation</td>
<td>Application form</td>
<td>Report</td>
</tr>
</tbody>
</table>
Data Seal of Approval + WDS = CoreTrustSeal

CoreTrustSeal offers to any interested data repository a core level certification based on the CoreTrustSeal Data Repositories Requirements.
CoreTrustSeal Catalogue of Requirements

Repository Audit and Certification Catalogues

The Challenge:
Creating harmonized Common Procedures for certification of repositories

What is the solution?
Drawing from the procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (WDS), the RDA Repository Audit and Certification DSA-WDS Partnership World Group has produced a convergent DSA-WDS certification standard aiming to eliminate duplication of effort, increase certification procedure coherence and comparability thus benefiting research, data managers, librarians and scientific communities.

Produced by: The Repository Audit and Certification DSA-WDS Partnership WG
https://rd-alliance.org/groups/rdaswds-certification-digital-repositories-wg.html

What is the impact?
The convergent DSA-WDS certification standard marks a step towards having more coherent, increasingly stringent and compatible standards for repository certification. The harmonization of repository certification criteria will increase adoption and the number of certified repositories building trust for data generators, data consumers and funding bodies.

Find out more about the Repository Audit and Certification DSA-WDS Partnership WG Recommendation

European Commission > Strategy > Digital Single Market >
Digital Single Market
Identification of ICT specifications
CORETRUSTSEAL DATA REPOSITORY CERTIFICATION

Promote trust and confidence in your shared data resources.

Apply for certification

DATA REPOSITORIES REQUIREMENTS
Explore the 16 Core Trustworthy Data Repositories requirements which are intended to reflect the characteristics of trustworthy repositories.

READ MORE →

HOW TO APPLY
We encourage repositories to seek core certification against Trustworthy Data Repositories Requirements

READ MORE →

LIST OF CERTIFIED REPOSITORIES
Explore CoreTrustSeal certified data repositories

READ MORE →

16+
Requirements

130+
Certified Repositories
CoreTrustSeal
Data Repository Certification

16 Requirements:

- Context (R0)
- Organizational infrastructure (R1-6)
- Digital object management (R8-14)
- Technology (R15-16)
- Applicant feedback

https://doi.org/10.17026/dans-22n-gk35
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
Two step certification process

1. **Self assessment** based on 16 Requirements (URLs of documented public evidence + compliance level) with extended guidance

2. **Peer review** by two expert and independent reviewers under the responsibility of the CoreTrustSeal Standards and Certification Board

- Successful applications are made publicly available
- Certification valid 3 years
- Administrative fee of 1,000 euro
- Certification mandatory for regular members of WDS
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).
Digital object management

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.
Digital object management

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:

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., fixity 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.
Self assessment

- Self assessment based on requirements
- Online tool available
- Extended guidance (document and webinar) available
- URLs to evidence strongly encouraged
- Maturity ratings strongly encouraged
- Responses in English
- Evidence documents in other languages need a short summary in English

- Review of the self assessment by two reviewers under the responsibility of the CTS Board
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 the institute with specific expertise temporarily involved in the actual work
Certification in practice: the 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 requirements

- The effort will rise when you climb the certification stairs to an extended and formal certification level
CTS uptake
Why do repositories invest in certification efforts?

- Builds stakeholder confidence in the repository (funders, host organizations, publishers, etc.)
- Raises awareness about digital preservation
- Improves communication within the repository
- Improves repository processes
- Ensures transparency
- Differentiates the repository from others
Why do we do it 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
Main takeaways

• Start with an internal quick scan to determine the level of entry against the catalogue of requirements
• Even without formal certification the requirements can be used as a benchmark opportunity, to identify gaps and areas that need more attention
• Commitment from the top is crucial
• Broad support within the organization is needed
• Do not aim too high at once; start with CTS certification
Important links

CoreTrustSeal website:
https://www.coretrustseal.org

WDS webpage on certification:
https://www.icsu-wds.org/services/certification

Survey of investments in and benefits of core certification:
Thank you for listening

ingrid.dillo@dans.knaw.nl
www.dans.knaw.nl