

Building Trust in Scientific Data: Certification & the CoreTrustSeal

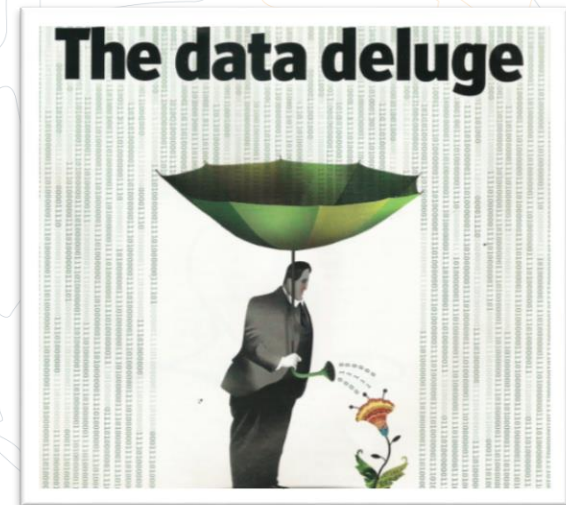
*International Workshop on Sharing,
Citation and Publication of Scientific Data
Across Disciplines*

Rorie Edmunds
ICSU-WDS Programme Officer



Data, Data Everywhere

- ‘Data Revolution’ has led to a growing recognition of the value of research data
- Witnessing strong trend of Open Science supported by data policies to ensure proper data sharing
- Societal and political pressure towards accountability, transparency and verifiability of science make data preservation and sharing part of scientific integrity
- Reuse and repurposing of data is recognized as a scientific and economical imperative, including by research funders, who are increasingly mandating science be ‘open’



Technical Limitations

Technical barriers to data sharing include:

- A system does not operate according to its objectives and specifications
- Datasets are not complete or include unintended modifications
- Datasets do not contain what they claim to contain
- Access to data & services is not guaranteed
- Datasets & services are not usable (for whatever reason)



WORLD DATA SYSTEM

Social Limitations: Trust

- Being unable to trust data from other sources is one of the cultural challenges preventing proper data management, preservation, and sharing (along with data ownership and the fear of being discredited or scooped)
- Trust is thus at the very heart of storing and sharing data:
 - Data funders want reassurances that their investment in the production of research data is not wasted, but will remain into the future; whilst data reuse data will give them a higher return on investment
 - Data depositors want to be sure their data are safe and remain accessible, usable, and meaningful over time
 - Data users want to know that data have been preserved properly and are of high quality



Trustworthy Data Repositories



- Certification standards play an important role in establishing trust, and hence sustaining the opportunities for long-term data sharing
- The demand for a way to evaluate the trustworthiness of data repositories is being answered, with a number of certifications services becoming available
- These standards not only take into account technical infrastructure and standards, but also look at organizational, financial, staffing, and legal aspects, as well as workflows, risk management, etc.



Certification Framework

A number of synergetic or complementary certification procedures now co-exist:

- **Core Certification:** The CoreTrustSeal, which came from the community-based norms granted to data repositories who obtained the DSA or WDS membership
- **Extended Certification:** Granted to repositories who perform a structured, externally reviewed and publicly available self-audit based on DIN 31644/nestor Seal
- **Formal Certification:** Highly demanding standard involving a full external audit and certification based on ISO 16363



Benefits of TDRs for a Data Producer/Depositor



- Helps fulfil your Data Management Plan (i.e., satisfy funders/open data requirements)
- Preserves your initial investment of collecting data
- Satisfaction that your data are being stewarded correctly and remain useful and meaningful
- Your data are looked after long term, even if the repository discontinues
- Increases the ease of discovery of your data.
- Facilitates the publication, reuse/repurposing, and citation of your data
- Recognized expertise is available to assist you with technicalities
- Ensures any conditions you want on access and use, as well as licensing, are adhered to

Benefits of TDRs for a Data Repository (Part 1)



- Improves the performance, agility, quality, and transparency of your processes
- Increases internal communication between different sections/groups of your repository
- Helps ensure your reliability and durability over a long period of time
- Improves internal awareness of and compliance with established standards
- Offers a benchmark for comparison and helps determine strengths and weaknesses

Benefits of TDRs for a Data Repository (Part 2)



- Displays your commitment to data and service quality and to long-term data curation
- Highlights your data holdings and services are easily searchable, rapidly accessible, and satisfy national and international standards
- Increases your visibility, thus enhancing discovery of your data holdings by disciplinary and interdisciplinary user bases
- Demonstrates to your research stakeholders that an independent authority has found you follow good practices, thus building confidence
- Improves your national and international recognition and reputation, particularly with funders

Benefits of TDRs for a Data User



- Discovering what you are looking for is easy
- Understand your access and usage rights is easy
- Existing data can be reused/repurposed, thus avoiding costly collection or production
- Others' results can be verified (and thus built on), thus accelerating scientific knowledge
- Peers can be cited, knowing that their data will still exist into the future
- Satisfaction that the data is original/uncorrupted due to provision of full provenance
- Data can be easily used through the inclusion of complete/appropriate metadata in an international/community standard
- Feedback can be given to the data producer/holder

The CoreTrustSeal

Data Seal of Approval Certification of Trusted Data Repositories



ICSU-WDS Certification of Regular Members



Research Data Alliance Repository Audit and Certification DSA-WDS Partnership WG



CoreTrustSeal Certification 101

- Core certification is a minimally intensive two-step process whereby a data repository supplies evidence that it is sustainable and trustworthy:
 1. Repository completes an internal self-assessment based on online application containing 16 Requirements
 2. Application is reviewed by two community peers under the responsibility of the CoreTrustSeal Standards and Certification Board, taking into account the repository's specific aims and context
- The process is iterative in that reviewers provide feedback (supported by the Board) until the application is approved. Successful applications are then made publicly available
- Renewal is every three years
- Evidence = URLs of documented (ideally public) evidence
- NB: Completing a self-assessment is very useful even without core certification as it enables appraisal of internal procedures



Core TDR Requirements (Part 1)

- Background information:
 - Context
- Organizational infrastructure:
 - Mission/scope
 - Licenses
 - Continuity of access
 - Confidentiality and ethics
 - Organizational infrastructure
 - Expert guidance



DOI 10.5281/zenodo.168411

25/08/2015

Common Requirements/V2.1



DSA–WDS Partnership Working Group Catalogue of Common Requirements

Introduction

Importance of Certification

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that the data they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.

Basic Certification and its Benefits

Nowadays certification standards are available at different levels, from a basic level to extended and formal levels. Even at the basic level, certification offers many benefits to a repository and its stakeholders.

Core TDR Requirements (Part 2)

- Digital object management:
 - Data integrity and authenticity
 - Appraisal
 - Documented storage procedures
 - Preservation plan
 - Data quality
 - Workflows
 - Data discovery and identification
 - Data reuse



DOI 10.5281/zenodo.168411

25/08/2015

Common Requirements/V2.1



DSA–WDS Partnership Working Group Catalogue of Common Requirements

Introduction

Importance of Certification

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that the data they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.

Basic Certification and its Benefits

Nowadays certification standards are available at different levels, from a basic level to extended and formal levels. Even at the basic level, certification offers many benefits to a repository and its stakeholders.

Core TDR Requirements (Part 3)

- Technology:
 - Technical infrastructure
 - Security
- Applicant feedback



DOI 10.5281/zenodo.168411

25/08/2015

Common Requirements/V2.1



DSA–WDS Partnership Working Group Catalogue of Common Requirements

Introduction

Importance of Certification

National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

If we want to be able to share data, we need to store them in a trustworthy digital repository. Data created and used by scientists should be managed, curated, and archived in such a way to preserve the initial investment in collecting them. Researchers must be certain that data held in archives remain useful and meaningful into the future. Funding authorities increasingly require continued access to data produced by the projects they fund, and have made this an important element in Data Management Plans. Indeed, some funders now stipulate that the data they fund must be deposited in a trustworthy repository.

Sustainability of repositories raises a number of challenging issues in different areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital repositories and hence the potential for sharing data over a long period of time. By becoming certified, repositories can demonstrate to both their users and their funders that an independent authority has evaluated them and endorsed their trustworthiness.


Basic Certification and its Benefits

Nowadays certification standards are available at different levels, from a basic level to extended and formal levels. Even at the basic level, certification offers many benefits to a repository and its stakeholders.

Example Requirement

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.

Compliance Level 

Response

Guidance:

Repositories must ensure that data can be understood and used effectively into the future despite changes in technology. This Requirement evaluates the measures taken to ensure that data are reusable.

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

- Which metadata are required by the repository when the data are provided (e.g., Dublin Core or content-oriented metadata)?
- Are data provided in formats used by the Designated Community? Which formats?
- Are measures taken to account for the possible evolution of formats?
- Are plans related to future migrations in place?
- How does the repository ensure understandability of the data?

Reuse is dependent on the applicable licenses covered in R2 (Licenses).



WDS Objectives

- Enable universal and equitable (full and open) access to quality-assured scientific data, data services, products, and information
- Ensure long-term data stewardship
- Foster compliance to agreed-upon data standards and conventions
- Provide mechanisms to facilitate and improve access to data and data products



WDS Accreditation



1. Complete an Expression of Interest:
www.icsu-wds.org
2. CoreTrustSeal Certification
3. Adhere to WDS Bylaws and sign Letter of Agreement
4. Regular membership granted by the WDS-SC.
5. Renewal every three years

Network Accreditation

- Networks (umbrella organizations) are currently outside the scope of the CoreTrustSeal, and will continue to be accredited solely by ICSU-WDS
- Networks vary greatly in their makeup and responsibilities (e.g., central bureau, coordinating committee, secretariat, programme office)
- As well as how they take responsibility for the competence and ongoing performance of the network components (e.g., data repositories, analysis centres, and products)
- In particular, the Network organization and aims, and how it takes responsibility for the quality of its component nodes is examined

WDS Network Member Accreditation Requirements



General Requirements & Policies

- Letter of Agreement with ICSU
- External experts to provide advice and guidance
- WDS biennial meetings
- Active communication with research community/users
- Full, open, timely, unrestricted access to data, metadata...

Organizational Framework

- Defined scope, responsibility for long-term preservation, target user community and needs, rights of users to access data, processes to respond to change
- Adequate in terms of funding, staff, long-term planning
- Scientific expertise: offers oversight of international reputation to nodes
- Continuity plan for nodes
- Committed to formal periodic review and assessment

Network Framework Description

- Systems to assess the capability of network components, and information about how new nodes are recruited, and if relevant, capacity building undertaken to improve the operational performance. Reporting of nodes joining/leaving.

The (Almost) Here and Now

- From January:
 - CoreTrustSeal will become an independent organization – the DSA will close and WDS will be a partner providing some initial administrative support
 - Fee to cover (only) administration costs will start on 1 January. A legal entity is being set up in the Netherlands
 - Interim Board will become the first official CoreTrustSeal Board until elections can take place
 - The Assembly of Reviewers will be formed from volunteers from the DSA and WDS communities
 - The CoreTrustSeal Application Management Tool under development will be released

The Future of CoreTrustSeal

- Create a global shared certification framework that includes other standards
- Provide certifications standards for repositories of other digital entities within Global Research Infrastructure:
 - Networks
 - Data-analysis services
 - Code/software
 - Ontologies/Vocabulary services
 - Images
 - ...
- It is not necessarily expected that the CoreTrustSeal would develop such standards, but rather manage them once consensus is reached with a community



Software Repositories

- Things that are slightly different:
 - R0 Level of curation – What does ‘enhanced’ mean in the case of software?
 - R1 Mission/Scope – More ‘non-research specific’ repositories used for software
 - R2 Licenses – What does it mean to maintain software licenses?
 - R6 Expert guidance – What about non-research specific repositories?
 - R8 Appraisal – Does one need different/additional policies (e.g., for understanding execution)?
 - R12 Workflows – Specific to software deposit, ‘timing’ may be different
- Things that require discussion
 - R11 Data quality – What should the assessment of ‘software quality’ comprise?
 - R14 Data reuse – More than metadata required for software reuse. What does format change mean for software?

More Information

www.CoreTrustSeal.org

The Hague | Tokyo | +31 6 2386 3243 | +81 4 2327 6395 | info@coretrustseal.org

CORE TRUST SEAL

Home About Certification Apply Contact

CORETRUSTSEAL CERTIFIED DATA REPOSITORIES

Broad disciplinary and geographic coverage

[Browse Map and List](#)

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](#)

info@coretrustseal.org

www.ICSU-WDS.org

Search

Home Strategy Login Contact Feedback

Trusted Data Services for Global Science

ICSU WORLD DATA SYSTEM

Home About Community Data and Services Publications News Events

ICSU World Data System DATA STEWARDSHIP AWARD

Winner of WDS Data Stewardship Award 2016: Boris Biskaborn

Latest News | **WDS Blog**

SciDataCon 2016 Programme Available + IDW Early Bird Registration Closing Soon!

01 Aug 2016

We are delighted to publish the draft programme for SciDataCon 2016 with options for viewing in a full window or for downloading the spreadsheet. The two days (12–13 September) will contain a total of 56 sessions spread over 7 parallel tracks, as well as 2 dedicated poster sessions. Details will be provided soon on keynotes and on the opening ceremony on the evening of Sunday, 11 September, as ...
[Read more](#)

Boris Biskaborn Wins 2016 WDS Data Stewardship Award

Upcoming Events & Deadlines

26 Aug 2016
Sheraton Denver Downtown Hotel
Room Block Expires for IDW 2016

10 Sep 2016
Sheraton Denver Downtown Hotel, 1550 Court Place, Denver, Colorado 80202, USA
ICSTI ITOC and TACC Workshops

11 to 13 Sep 2016
Sheraton Downtown Hotel in Denver, Colorado
SciDataCon 2016 – Advancing the Frontiers of Data in Research

[Full Listing](#) [Calendar](#) [Add Event](#)

IPO@icsu-wds.org



INTERNATIONAL DATA WEEK IDW 2018

Gaborone, Botswana: 05–08 November 2018

