Developments in the certification of data centres, services and repositories through an RDA/WDS/DSA partnership

Lesley Rickards, Mary Vardigan, Ingrid Dillo, Françoise Genova, Hervé L'Hours, Jean-Bernard Minster, Rorie Edmunds, Mustapha Mokrane

Overview

- ICSU World Data System

- Membership of WDS

- Certifications including the Data Seal of Approval

- Research Data Alliance

- Certification WG tasks

- WG Outcome and Recommendations

- Implementation

- Conclusions

ICSU World Data System (ICSU-WDS)

- ICSU-WDS builds on the 50+ year legacy of WDCs and FAGS
  - Set up for International Geophysical Year (1957–1958)
  - Long-term stewardship, curation, archiving, and dissemination of scientific research data
- Experience from International Polar Year (2007–2008)
  - A new approach needed
- ICSU General Assembly (2008)
  - Replaced WDCs and FAGS by ICSU World Data System
- "New system vows for a better coordination and disciplinary coverage to respond efficiently to the needs of the new scientific research challenges under the ICSU umbrella"

ICSU-WDS Goals

- Enable universal and equitable access to quality-assured scientific 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 Membership (October 2016)

- **65 Regular:** organizations that are data stewards and/or data analysis services

- **10 Network:** umbrella bodies representing groups of data stewardship organizations and/or data analysis services

- **6 Partner:** organizations that are not data stewards or data analysis services, but that contribute support or funding

- **18 Associate:** organizations interested in the WDS endeavour, but that do not contribute direct funding or other support

Practicalities to be considered

- Keep it as simple as possible

- Maintain transparency

- All criteria are mandatory, but there is flexibility

- What counts as a 'pass'? Different levels?

- Who does what in managing the task and assessing the completed applications – i.e. roles for WDS-SC, WDS IPO, CODATA, Scientific Unions, Others...

- Downloadable version of application with guidance is available

WDS Certification Procedure

- Facility responded to initial WDS survey, or provides a letter of interest

- Facility demonstrates its capabilities using the on-line application to describe its capabilities

- If necessary, an on-site review may take place (to be decided by negotiations with the candidate)

- Accreditation as a WDS member

- Review of accreditation should take place approx. every 5 years

Metrics

- Goal is to have objective controls (criteria) against which candidate WDS members can be evaluated

- Some questions can be answered as yes or no

- For many, evaluation based on non-standardized information

- Information supplied for a specific criterion can be attributed to different levels of maturity, e.g.
  - not addressed
  - conception phase
  - implementation phase
  - operational

Criteria for WDS Certification

Four categories (for Regular Members)

- WDS general requirements and policies (Organisation specific requirements) (5)

- Organisational framework (5)

- Management of data, products and services (4)

- Technical infrastructure (3)

Summary details available on the WDS web-site

Open for applications since February 2011

Why have Trustworthy Digital Repositories?

- Data created and used by scientists should be managed, curated, archived, and made accessible
- Researchers (and others) must be certain that data held in repositories remain useful and meaningful into the future
- Funders increasingly likely to mandate open data and data management policies
- Science publishers may stipulate data must be deposited in a trustworthy repository

Data repositories / centres…

- Mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future
- Constant monitoring, planning, and maintenance

Certifications schemes...

- Reference Model for an Open Archival Information System (OAIS)

- Trustworthy Repositories: Audit & Certification (TRAC): Criteria and Checklist

- Catalogue of Criteria for Trusted Digital Repositories, Nestor materials

- Data Seal of Approval

- ISO 16363

- WMO Information System

- UK Marine Environmental Data and Information Network

- IOC/IODE Quality Management Framework

What is the RDA?

- **Research Data Alliance (RDA) planned / launched in 2013**
  - European Commission,
  - US National Science Foundation
  - Australian Government (through ANDS)

- **To address the growing global need for data infrastructure**

- **An international, community-powered organization**

- **RDA's vision:** "researchers / innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society"

- **Experts work together: self-forming / voluntary manner**
  - Focused Working Groups
  - Exploratory Interest Groups
  - Birds of a Feather Groups

- **Further information: www.rd-alliance.org**

RDA/DSA/WDS WG Starting Point

- Data Seal of Approval (DSA) and World Data System (WDS) came together in a WG under the auspices of the RDA Interest Group on Repository Audit and Certification

- Goal was to harmonize core certification requirements and procedures, ultimately setting the stage for a global shared framework including other standards

> **Broader goal:** to inspire trust, which is at the heart of sharing and archiving data.
> Trust among: Users, Depositors, Funders

- Understand threats to and risks within its systems

- Encourage regular cycle of audit and/or certification

The Two Partners

ICSU World Data System

- Originally Earth and Space Sciences
- Assessment to allow membership
- 17 criteria
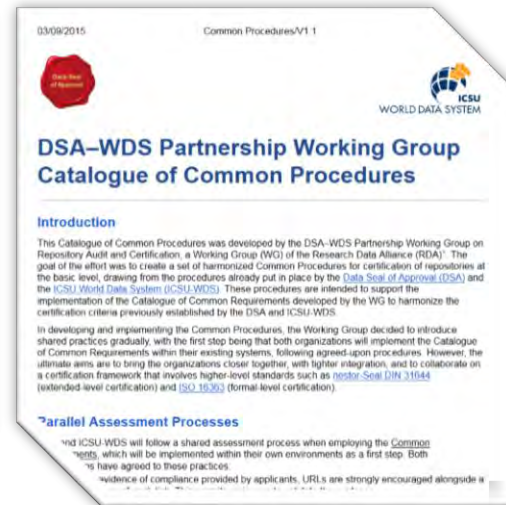- Review every 5 years
- 65 regular members; more underway

Data Seal of Approval

- Originally Humanities and Social Sciences
- 16 guidelines for trustworthy digital repositories (data producers, repositories, consumers)
- DSA granted for a period of 2 years
- 62 seals acquired; more underway

What was the plan?

- Common Requirements – Basic certification criteria

- Common Procedures – Implementation plan for introducing requirements in partnership

- Testbed – "Real-world" valuation of Common Requirements and Procedures

Common Requirements

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.

http://tinyurl.com/pm9sflp

Structure of Common Requirements

- Introduction
  - Importance of certification
  - Basic certification and its benefits
- Background and General Guidance
  - Goals and context
  - Suggested compliance/maturity levels
- Governance
- Glossary

DSA/WDS Requirements Mapping

- Mapped DSA to WDS and WDS to DSA

- Derived mappings along with notes on level of the match (good match, partial, gap, etc.)

- Reconciled the two standards with suggested common language for requirements

- Assigned a concept to each common requirement, e.g., Discovery, Appraisal, Continuity of Access

- Assigned ISO/TRAC label(s): Organisational Infrastructure, Digital Object Management, Technology

Issues and Problems: Mapping DSA and WDS

Partial Match

Common Requirement	DSA Guideline	WDS Criterion
The repository maintains all **applicable licenses covering data access and use** and monitors compliance.	14. The data consumer complies with access regulations set by the data repository. 16. The data consumer respects the applicable licences of the data repository regarding the use of the data.	[16.4] The facility has defined: the rights of its users to access and use data. IV. Organisational framework

Gap

Common Requirement	DSA Guideline	WDS Criterion
The repository adopts mechanism(s) to secure **ongoing scientific guidance and feedback from recognized experts**, and maintains publicly accessible documentation of such guidance.		12. Have relevant external experts to provide advice and guidance to WDS node. III. General requirements.

Issues and Problems: Mapping DSA and WDS

Poor Match

Common Requirement	DSA Guideline	WDS Criterion
The repository has a **continuity plan** to ensure ongoing access to and preservation of its holdings.	9. The data repository assumes responsibility from the data producers for access and availability of the digital objects.	19. Maintenance of a continuity plan in the event of a host institution shift of interests or reaction to substantial changes. IV. Organisational framework

Good Match

Common Requirement	DSA Guideline	WDS Criterion
The repository applies **documented processes and procedures** in **managing archival storage** of the data.	6. The data repository applies documented processes and procedures for managing data storage.	23. Archival storage of the data sets is undertaken to defined specifications. V. Management of data, products, and services

Issues, problems, solutions

- Lists have similarities and differences

- DSA guidelines more concise; WDS has multi- part criteria

- DSA focus is on data management, not organisational stability

- WDS certification includes membership in the WDS and certification of services, not in scope for the DSA

- Lengthy discussions on each guideline, attempting to separate WDS focus on membership and services

- Overall, working together beneficial

What has been achieved?

16 Common Requirements

- **Organisational infrastructure**

Context	Mission / Scope	Licences
Continuity of access	Confidentiality / ethics	
Expert guidance	Organisational infrastructure	

- **Digital object management**

Data integrity and authenticity		
Appraisal	Documented storage procedures	
Preservation plan	Data quality	Workflows
Data discovery and identification		Data re-use

- **Technology**

Technical infrastructure	Security

http://tinyurl.com/pm9sflp

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.

Certification and its Benefits

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

Common Procedures

03/09/2015   Common Procedures/V1.1

DSA–WDS Partnership Working Group Catalogue of Common Procedures

Introduction

This Catalogue of Common Procedures was developed by the DSA–WDS Partnership Working Group on Repository Audit and Certification, a Working Group (WG) of the Research Data Alliance (RDA)[1]. The goal of the effort was to create a set of harmonized Common Procedures for certification of repositories at the basic level, drawing from the procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS). These procedures are intended to support the implementation of the Catalogue of Common Requirements developed by the WG to harmonize the certification criteria previously established by the DSA and ICSU-WDS.

In developing and implementing the Common Procedures, the Working Group decided to introduce shared practices gradually, with the first step being that both organizations will implement the Catalogue of Common Requirements within their existing systems, following agreed-upon procedures. However, the ultimate aims are to bring the organizations closer together, with tighter integration, and to collaborate on a certification framework that involves higher-level standards such as nestor-Seal DIN 31644 (extended-level certification) and ISO 16363 (formal-level certification).

Parallel Assessment Processes

DSA and ICSU-WDS will follow a shared assessment process when employing the Common Requirements, which will be implemented within their own environments as a first step. Both organizations have agreed to these practices:

- In the evidence of compliance provided by applicants, URLs are strongly encouraged alongside a description of each link. This permits reviewers to validate the evidence.

http://tinyurl.com/os6vb94

What has been achieved?

Common procedures

Implementation – introducing requirements in partnership

- URLs to evidence strongly encouraged
- Maturity ratings strongly encouraged
- Assessments to be publicly available
- Renewals every three years
- Three entry points for certification:
  (DSA, WDS, and new website)
- Successful completion means certification in both DSA and WDS
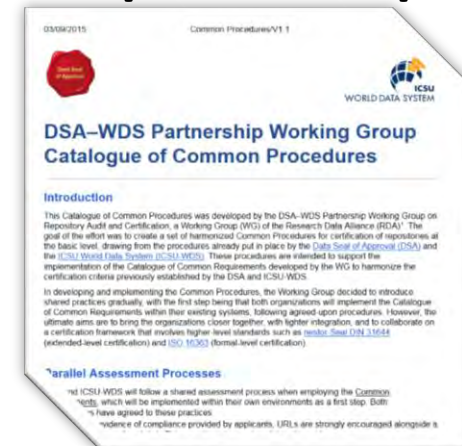
Sustainable Review Process

- Pool of reviewers (training provided)
- Approved by the new DSA–WDS Certification Board

Mutual Governance Process

http://tinyurl.com/os6vb94

What has been achieved?

- **Testbed – "Real-world" evaluation of Common Requirements and Procedures**

- *Develop common testbed, and surrounding organizational framework, for peer review and certification... [to] provide practical insight into the proposed common WDS–DSA catalogue and review process, thus enabling iterative improvements to those procedures*

- Tested by 6 organisations
  - Diverse set of repositories and services represented
  - Procedures will follow agreed processes as much as possible
- Comments from testers feeding in to improving Common Requirements and Common Procedures
- Testbed evaluation report

Testbed evaluation report

<u>Positive comments:</u>

" …clear and understandable"

" …structure seems to be more suitable"

"…stronger emphasis on documented procedures and plans"

"…easier to meet the common requirements but …some need a more detailed description and evidence."

<u>Further work needed:</u>

- clarify the cross-disciplinary nature of the requirements - explaining ethical considerations
- further define outsourcing services
- clarify the subtle differences between the task of evaluating technical quality and the issue of research quality
- further explain data management and OAIS repository concepts
- repetition of requests (e.g. for metadata information)
- guidance on maximum length of answer

What next?

- Working Group has completed its tasks - Interest Group to continue work
  - Consider Data Services and Networks
  - Conduct outreach to publishers
  - Explore linkages with **re3data.org**
- DSA and WDS formally adopt and implement Core Common Requirements
- Resolve "Needed" items in the Common Procedures document
- Conduct outreach to other standards like nestorSEAL and ISO

Towards a global certification framework

Conclusion - What is the impact?

- Will provide a step towards having more coherent, increasingly stringent and compatible standards for repository certification

- DSA–WDS certification standard adoption will create a critical mass of certified repositories across a range of domains and disciplines

- Data Collectors, Funders, Publishers and Users – deliverable inspires trust, which is at the heart of sharing and archiving data