Data Archiving and Networked Services

DANS

# New Common Requirements for Core Level Certification of Trustworthy Digital Repositories: DSA and WDS working together

Data Seal of Approval

Ingrid Dillo, DANS, The Netherlands

IASSIST2016, Session 1E Policies and Trust
Bergen, 3 June 2016

ICSU
WORLD DATA SYSTEM

# Data Seal of Approval

## Overview

Ingrid Dillo, DANS, The Netherlands

IASSIST 2015, Session Trust me!

Minneapolis, 5 June 2015

Data Seal of Approval

# Looking Ahead

- In process of activating the DSA General Assembly to bring on more peer reviewers and ensure sustainability of the effort

- Working with World Data System on harmonizing guidelines and procedures between the two organizations

"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?"

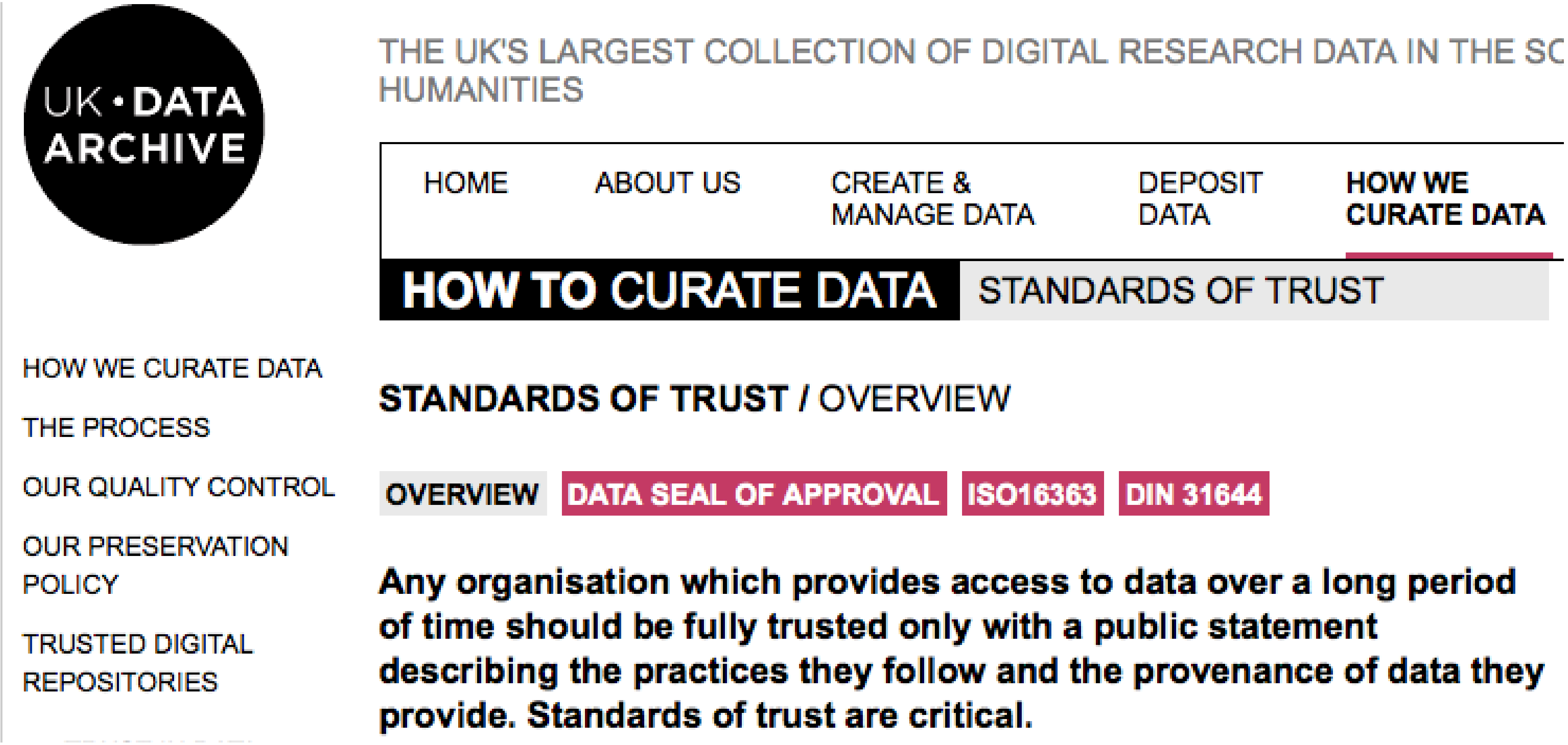
The Data Harvest:

How sharing research data can yield knowledge, jobs and growth

An RDA Europe Report

December 2014

# Trust in data archives: an example



THE UK'S LARGEST COLLECTION OF DIGITAL RESEARCH DATA IN THE SC HUMANITIES

**STANDARDS OF TRUST /** OVERVIEW

**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

ISO

nestor Seal 2016

Data Seal of Approval

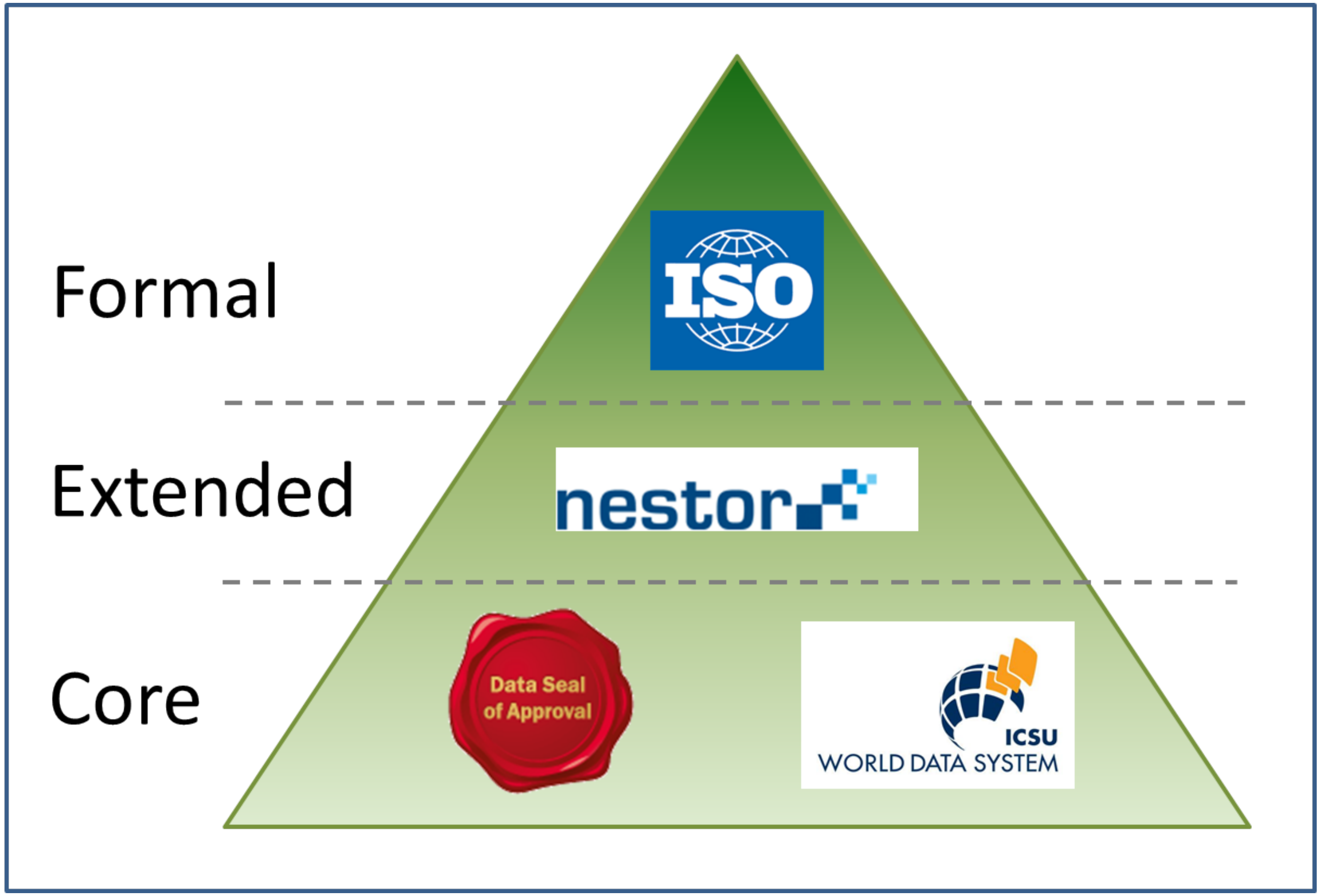
# Tiered European Framework

**Core Certification** is granted to repositories which obtain DSA certification

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

**Formal Certification** is granted to repositories which obtain full external audit and certification based on ISO16363

# Global Certification Framework



# DSA key characteristics

- 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)
- Some 60 seals acquired, some 50 in progress

# WDS key characteristics

ICSU WORLD DATA SYSTEM

- World Data System part of ICSU
- Light-weight certification procedure for regular and network members
- Based on self assessment
- Peer review by WDS Scientific Committee
- Focus on earth and spatial sciences
- Many members in US and Asia
- Renewal between 3 and 5 years
- Some 70 accredited members

# Core level certification: DSA and WDS

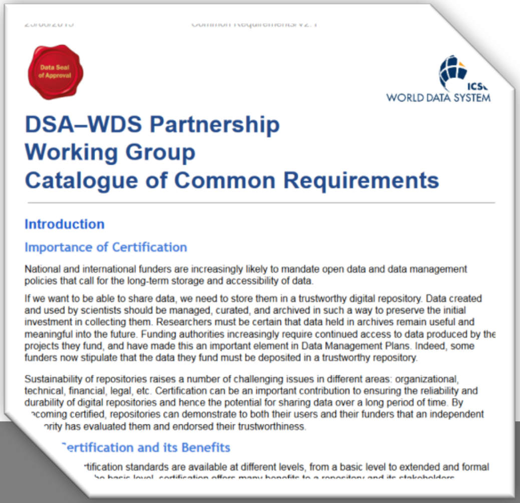
- Data Seal of Approval and World Data System both basic, lightweight mechanisms for repository assessment with many commonalities

- DSA and WDS came together in a WG under the auspices of the RDA Interest Group on Repository Audit and Certification

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

# Working Group Outcomes

- Common Requirements – Basic certification criteria
- Common Procedures – Implementation plan for introducing requirements in partnership
- Testbed – "Real-world" valuation of Common Requirements and Procedures



23/06/2015 Common Requirements/v2.1

Data Seal of Approval

ICSU WORLD DATA SYSTEM

**DSA–WDS Partnership Working Group Catalogue of Common Requirements**

**Introduction**

**Importance of Certification**

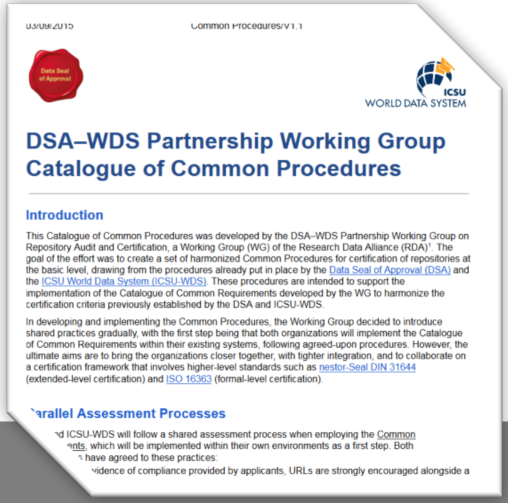
National and international funders are increasingly likely to mandate open data and data management policies that call for the long-term storage and accessibility of data.

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 ecoming certified, repositories can demonstrate to both their users and their funders that an independent ority has evaluated them and endorsed their trustworthiness.

**ertification and its Benefits**

tification standards are available at different levels, from a basic level to extended and formal he basic level, certification offers many benefits to a repository and its stakeholders



03/09/2015 Common Procedures/V1.1

Data Seal of Approval

ICSU WORLD DATA SYSTEM

**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).

**arallel Assessment Processes**

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

vidence of compliance provided by applicants, URLs are strongly encouraged alongside a

# Common Requirements

- Background information:
  - Context

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

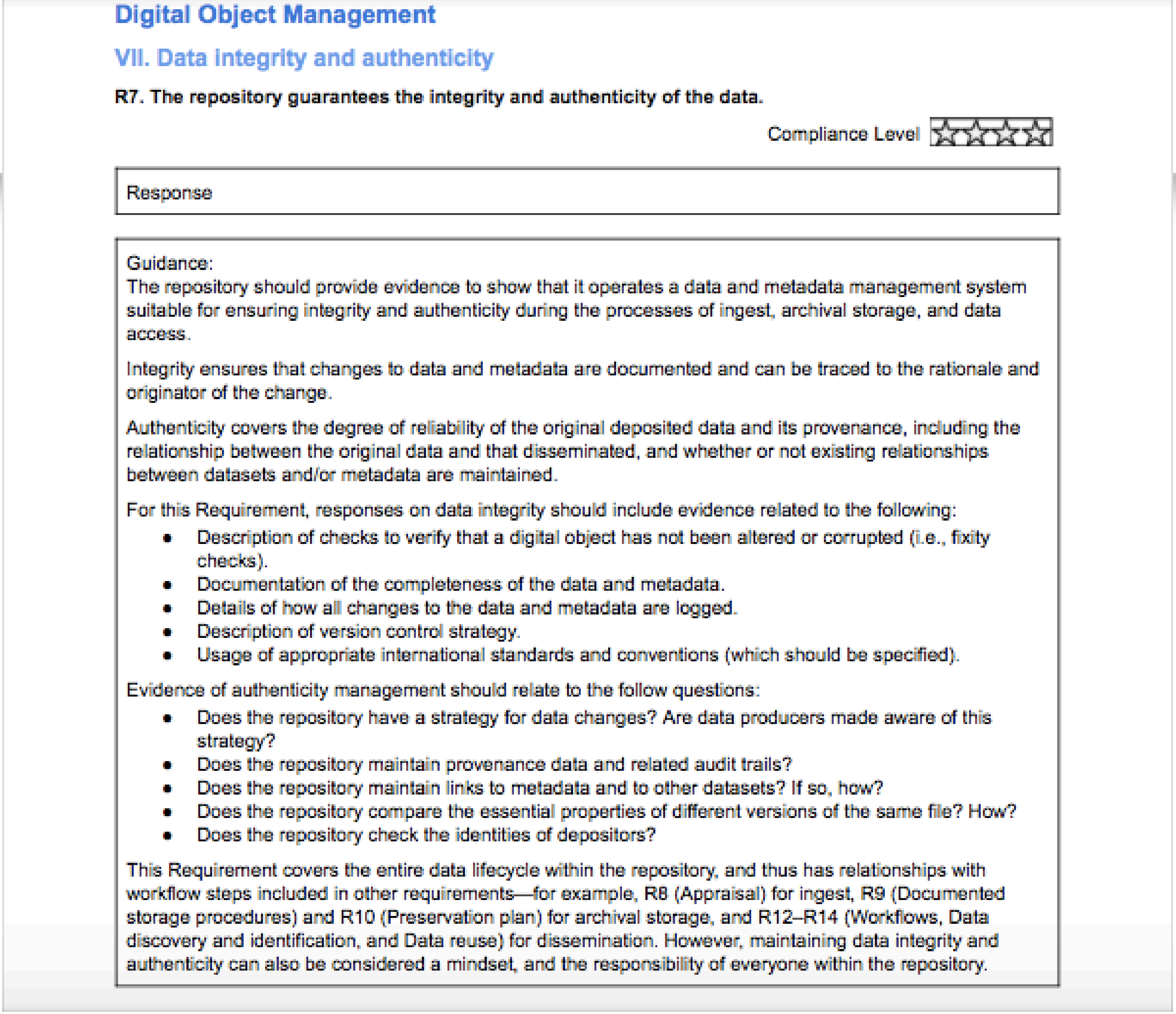
# Common Requirements

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

# Common Requirements

- Technology:
  - Technical infrastructure
  - Security

- Additional information and applicant feedback



## 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 follow 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.

# Common procedures

Parallel Assessment Processes:

- URLs to evidence strongly encouraged
- Maturity ratings strongly encouraged
- Assessments to be publicly available
- Successful completion means certification in both DSA and WDS
- 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

# The future of core certification

- Both DSA and WDS will introduce the new common requirements over the course of this year (launch at IDW)
- Continue work on the Common Procedures
- Conduct outreach to other standards like nestorSEAL and ISO proposal

# Benefits and Value of Core Certification (as noted by repositories)

- Builds stakeholder confidence in the repository
- Improves communication within the repository
- Improves repository processes
- Ensures transparency
- Differentiates the repository from others
- Saves time and labor over other certification methods

# Thank you very much for listening!

THIS WORLD DOES NOT NEED MORE TRUST
IT NEEDS MORE TRUSTWORTHINESS
made on imgur

ingrid.dillo@dans.knaw.nl

www.dans.knaw.nl

http://datasealofapproval.org/en/

https://www.icsu-wds.org/services/certification

https://rd-alliance.org/group/repository-audit-and-certification-dsa–wds-partnership-wg/outcomes/dsa-wds-partership