International Standards for Trustworthy Data Repositories

Robert R. Downs[1]

[1] *rdowns@ciesin.columbia.edu*

NASA Socioeconomic Data and Applications Center (SEDAC)

Center for International Earth Science Information Network (CIESIN)

The Earth Institute, Columbia University

Trustworthy Data Repositories Workshop
National Institutes of Health

Session 1: Trust Concepts and Standards
Monday, April 8, 2019; 9:00 a.m. - 10:30 a.m.

What are international standards for trustworthy data repositories?

- Development
  - Based on international standards
  - Focus on improving trustworthy data repositories
- Review
  - Evaluated as instruments to assess trustworthy data repositories
- Recognition
  - Informing practices of international data repositories
- Adoption
  - Used by research data communities to assess data repositories

CORE TRUST SEAL

Why are international standards needed to assess trustworthy data repositories?

- Standards establish expectations for performance
  - Recognized benchmarks can be compared and improved
  - International standards benefit from multinational activities
  - International communities can contribute to and benefit from standards
- Data repositories serve international communities
  - Online data repository use extends beyond national borders
  - Data collections often represent international observations
- Science is international
  - Research disciplines reflect multinational efforts and teams
  - Like scientific journals, repositories often serve disciplines
  - Global adoption of open science practices

Why assess or audit data repositories?

- **Data producers** need to know where to deposit their data
  - Trust that their data will be preserved, curated, and disseminated
- **Data users** need to know where they can find data
  - Data that are vetted, described for use, and available in the future
- **Funders** need to know who to support for data management
  - Where services are reviewed routinely for continuous improvement
- **Publishers** need to know who to recommend for archiving data
  - Where referenced data will be persistently accessible and usable
- **Data professionals** need to know where they can practice
  - Apply their data management skills
  - Obtain professional development in data stewardship
- **Data centers** need to know how they are performing
  - Policies, procedures, and practices that need improvement

Based on Downs, 2016.

Data Repository Audit: Cost vs Benefits

- Potential costs
  - **Time** for repository management and staff training
  - **Training** registration fees and travel costs, if applicable
  - **Time** for repository management and staff to prepare and be audited
  - Pre-audit improvements (**staff, software, hardware**), if applicable
  - Audit instrument **purchase fee**, if applicable
  - Audit **service fee**, including auditor travel costs, if applicable
- Potential benefits
  - Improve **transparency** and **quality assurance**
  - **Compare capabilities and services** with standard practices
  - **Review and identify gaps** in current and planned services
  - **Plan for needed enhancements** to improve capabilities and services
  - Obtain **certification** that recognizes attainment of standard, if applicable

CORE TRUST SEAL

Source: Downs, 2016.

Plan the Data Repository Assessment or Audit

- Planning by repository management
  - Identify and evaluate candidate audit instruments
  - Select the assessment instrument to be used
- Preparation by repository management and staff
  - Conduct a self-assessment using the selected instrument
  - Complete improvements to address identified weaknesses
- Scheduling formal assessment
  - Identify availability of auditors
  - Identify availability of repository managers and key staff

CORE TRUST SEAL

Based on Downs, 2016.

Compare Data Repository Assessment Instruments

- What does the candidate assessment instrument measure?
  - **Is it applicable** to the repository, its services, and capabilities?
  - Is the scope of the instrument consistent with **repository goals**?
  - **Do the metrics measure** whether each requirement has been satisfied?
  - Can it be **used internally** for pre-assessment or post-assessment reviews
- Cost of Data Repository Assessment
  - Many assessment instruments are available **free of charge**
  - **Costs vary** for performing formal data repository assessments
- Instrument Validity
  - Has the instrument been **developed** by a reputable organization?
  - Has the instrument been **reviewed** recently?
  - Is the instrument a **recognized stan**dard for assessing data centers?
  - Has it been **endorsed** by the community and by independent bodies?
  - Is the instrument actively **used** to audit and certify data repositories?

Based on Downs, 2016.

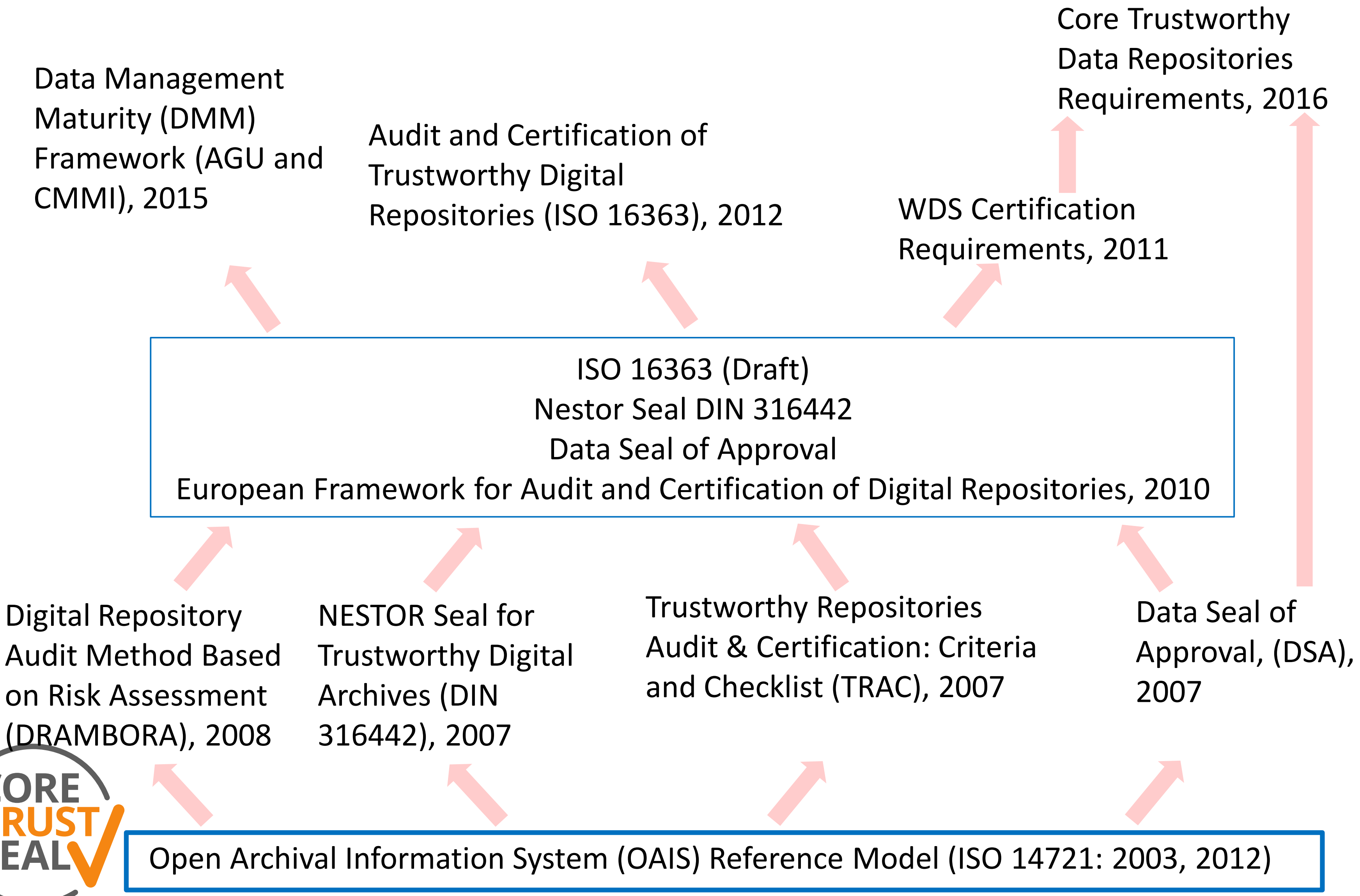
Data Repository Audit Considerations

		Authority	
		Internal	External
Frequency	Once	Internal one-time self-assessment	External one-time audit
	Periodic	Internal periodic self-assessments	External periodic audits

- Scope of audit:
  - Holistic vs targeted to specific capabilities, functions, or collections
- Approach should be based on objectives for the assessment
  - Why is the repository seeking an audit?
  - Which stakeholders are encouraging the audit?
  - Is improvement the primary objective or is the goal a credential?

Source: Downs, 2016.

Trustworthy Data Repository Certification Standards: An Evolutionary Perspective



Core Trustworthy Data Repositories Requirements, 2016

Data Management Maturity (DMM) Framework (AGU and CMMI), 2015

Audit and Certification of Trustworthy Digital Repositories (ISO 16363), 2012

WDS Certification Requirements, 2011

ISO 16363 (Draft)
Nestor Seal DIN 316442
Data Seal of Approval
European Framework for Audit and Certification of Digital Repositories, 2010

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), 2008

NESTOR Seal for Trustworthy Digital Archives (DIN 316442), 2007

Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), 2007

Data Seal of Approval, (DSA), 2007

Open Archival Information System (OAIS) Reference Model (ISO 14721: 2003, 2012)

Source: Downs, 2019

Selected Current Instruments for Assessing Scientific Data Repositories

- Audit and Certification of Trustworthy Digital Repositories (ISO 16363:2012)
  - Available free from CCSDS: https://public.ccsds.org/Pubs/652x0m1.pdf
- Core Trustworthy Data Repositories Requirements.
  - https://www.coretrustseal.org
- Data Management Maturity (DMM) Framework (AGU & CMMI)
  - http://dataservices.agu.org/dmm/
- Data Seal of Approval (DSA) http://www.datasealofapproval.org/en/
- Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)
  - https://www.repositoryaudit.eu/
- NESTOR Catalogue of Criteria for Trusted Digital Repositories (DIN 316442)
  - http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html
- Trustworthy Repositories: Audit & Certification (TRAC)
  - https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf

Based on Downs, 2016.

Current ISO Standards for Trustworthy Data Repositories

- ISO 14721: 2012, (CCSDS 650.0-M-2) Space data and information transfer systems -- Open archival information system (OAIS) -- Reference model. Free https://public.ccsds.org/Pubs/650x0m2.pdf

- ISO 16363: 2011, (CCSDS 652.0-R-1) Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. Free https://public.ccsds.org/Pubs/652x0m1.pdf

- ISO 16919: 2012, (CCSDS 652.1-M.2) Space data and information transfer systems -- Requirements for bodies providing audit and certification of candidate trustworthy digital repositories. Free https://public.ccsds.org/Pubs/652x1m2.pdf

Based on Downs, 2017.

The OAIS Reference Model

- ISO 14721, Reference Model for an Open Archival Information System (OAIS)
- Framework for repositories to provide long-term stewardship of data and other digital information
- Used as a basis for assessment instruments and activities
- Developed by Consultative Committee for Space Data Systems (CCSDS) in 2003
- Published by ISO in 2003, reviewed and revised in 2012
- Freely accessible: https://public.ccsds.org/Pubs/650x0m2.pdf
- ISO 14721:2012 **5-year review started in August 2016**

Based on Hughes & Downs, 2016.

OAIS Mandatory Responsibilities*

- **Negotiates for and Accepts Information**
  - Negotiate for and accept appropriate information from information Producers.
- **Obtains Sufficient Control**
  - Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.
- **Determines Designated Community**
  - Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.
- **Ensures Information is Independently Understandable**
  - Ensure that the information to be preserved is Independently Understandable to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information.
- **Follows Established Preservation Policies and Procedures**
  - Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed as part of an approved strategy. There should be no ad-hoc deletions.
- **Makes the Information Available**
  - Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity.

* Based on CCSDS. 2012. Reference Model for an OAIS.

ISO 16363

- ISO 16363:2012 Space Data and Information Transfer Systems – Audit and Certification of Trustworthy Digital Repositories
  - Published by the International Organization for Standardization (ISO)
- Developed by the Consultative Committee for Space Data Systems as CCSDS 652.0-M-1
  - Under review by the CCSDS Data Archive Interoperability (DAI) WG
  - Being reviewed in conjunction with review of the Open Archival Information System (OAIS) Reference Model (ISO 14721:2012)
  - Proposed revisions will be reviewed simultaneously by CCSDS and ISO

Freely available from ccsds.org:

https://public.ccsds.org/Pubs/652x0m1.pdf

CORE TRUST SEAL

Source: Hughes & Downs, 2016.

ISO 16363 Development

- Initiated in 2007 within CCSDS
  - Repository Audit and Certification (RAC) Working Group
  - Need criteria to assess OAIS compliance and guidance to reduce risk
  - Need for an international standard for assessing digital repositories
- Reference documents:
  - Open Archival Information System (OAIS) Reference Model (ISO 14721:2003)
  - Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
  - Catalogue of Criteria for Trusted Digital Repositories (Nestor Working Group)
  - Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)
  - OECD Guidelines for the Security of Information Systems and Networks
- Reviews
  - CCSDS and ISO communities' comments received and revisions applied
  - Test Audits Conducted at 6 repositories (3 in Europe, 3 in US)
- ISO 16363:2012 Published

Based on Downs, 2016.

Organization of ISO 16363

- **Organizational Infrastructure**
  - Governance and Organizational Viability
  - Organizational Structure and Staffing
  - Procedural Accountability and Preservation Policy Framework
  - Financial Sustainability
  - Contracts, Licenses, and Liabilities
- **Digital Object Management**
  - Ingest: Acquisition of Content
  - Ingest: Creation of the AIP
  - Preservation Planning
  - AIP Preservation
  - Information Management
  - Access Management
- **Infrastructure and Security Risk Management**
  - Technical Infrastructure Risk Management
  - Security Risk Management

Based on: Consultative Committee for Space Data Systems (2011) Audit and Certification of Trustworthy Digital Repositories: Recommended Practice. Magenta Book, Issue 1. Available: http://public.ccsds.org/publications/archive/652x0m1.pdf

Adoption and Use of ISO 16363:2012

- Endorsement by the Society of American Archivists Council
    - August 6, 2012
- Used for self-assessments and preparation
    - Data centers, institutional repositories, government agencies
- Used by professional development services
    - Training courses, workshops, presentations, consulting, and guidance
    - Audience: data creators, curators, repository managers, funders, consultants
- Used for Audits by PTAB
    - Conducted test audits of 6 repositories using draft ISO 16363 (2011)
    - Accredited for ISO 16363:2012 audit and certification (2017)
    - Offers training, conducts audits, reviews applications, and answers inquiries
    - National Cultural AudioVisual Archives (NCAA) ISO 16363 certified (2018)
    - U.S. Government Publishing Office ISO 16363 certified (2018)

Based on Downs, 2017.

ISO 16363 – CoreTrustSeal Relationship

- Complementary messages to increase awareness of instruments
  - Informing diverse communities on requirements for trustworthy repositories
- Mutually-informed development of instruments
  - Both instruments based on OAIS framework
- Self-Assessments for Repository Preparation
  - ISO 16363 Self-Assessment to prepare for CoreTrustSeal Certification
  - CoreTrustSeal Certification to prepare for ISO 16363 Audit
- Shared pathway for improving repository practices
  - CoreTrustSeal Certification -> ISO ISO16363 Certification
- Improvement of certification processes
  - Experiences conducting audits can inform auditing practices
- Improvement of Requirements
  - Experiences with audit instruments can inform improvement of instruments

Based on Downs, 2017.

Adopting Data Repository Assessment Instruments

- Select the assessment instrument to be adopted
  - Review candidate instruments and their use by others
- Obtain a free repository assessment instrument:
  - https://public.ccsds.org/Pubs/652x0m1.pdf
  - https://www.coretrustseal.org/why-certification/requirements/
- Review each requirement
  - Identify staff with expertise for each requirement
- Conduct a self-assessment
  - Identify evidence for meeting each requirement
  - Document evidence to describe how requirement is met

Based on Downs, 2016.

Summary: International Resources for Trustworthy Data Repository Assessment

- Standard framework for data repository trustworthiness
  - **Reference Model for an Open Archival Information System (OAIS).** (ISO 14721:2012, Consultative Committee for Space Data Systems. CCSDS.650.0-M-2) https://public.ccsds.org/Pubs/650x0m2.pdf
- Selected data repository assessment instruments
  - Audit and Certification of Trustworthy Digital Repositories. (ISO 16363, Consultative Committee for Space Data Systems. CCSDS.0-M-1) https://public.ccsds.org/Pubs/652x0m1.pdf
  - Core Trustworthy Data Repositories Requirements. CoreTrustSeal https://www.coretrustseal.org/
  - Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)
  - https://www.repositoryaudit.eu/
  - NESTOR Seal for Trustworthy Digital Archives. Deutsches Institut für Normung (DIN) 31644 Information and Documentation – Criteria for Trustworthy Digital Archives. http://www.dnb.de/Subsites/nestor/EN/Siegel/siegel.html

Based on Downs, 2018.