Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Audit of a Scientific Data Center for Certification as a Trustworthy Digital Repository: A Case Study

Robert R. Downs and Robert S. Chen

rdowns@ciesin.columbia.edu, *bchen@ciesin.columbia.edu*

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
The Earth Institute, Columbia University

2011 FALL AGU MEETING
San Francisco, CA
IN52A: Issues in Scientific Data Preservation and Stewardship
December 9, 2011

Scientific Data Stewardship Services

- Enable discovery, access, and use by current and future communities of users
  - Scientists from discipline
  - Scientists, professors, and students from various disciplines
  - Planners, policy makers and decision makers
- Manage and preserve data to foster science
  - Integrated data products and services for new analyses
  - Replication studies and longitudinal studies

Improving Data Stewardship

- Establishment of policies and procedures for managing and preserving data
- Professional development of data managers
- Adoption of data management systems and applications
- Procurement of new hardware and software
- Development of services that provide new capabilities for users and staff
- Evaluation of services for compliance with current standards and practices

Why improve data stewardship?

- Ensure that data will be usable in future
- Enhance capabilities for preservation, access, and use
- Reduce data stewardship costs
- Obtain trust of data depositors, users, and sponsors

Why certify repositories as trustworthy?

- Producers need to know whether a repository will be able to preserve and disseminate their data for use in the future.
- Users need to know which repositories can be relied upon to provide data that are authentic.
- Sponsors need to know which repositories to recommend for storing data produced by projects.
- Repositories need to know whether they are meeting current requirements for data stewardship.

Standards for Trustworthy Repositories

ISO 16363 Space Data for Information Transfer Systems – Audit and Certification of Trustworthy Digital Repositories, developed by the Consultative Committee for Space Data Systems (CCSDS)

ISO 16919 Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories, developed by the Consultative Committee for Space Data Systems (CCSDS)

Trusted Repositories Audit and Certification: Criteria and Checklist (TRAC), developed by the National Archives and Records Administration (NARA) and the Research Libraries Group (RLG)

ISO 14721:2003 Space data and information transfer Systems - Open archival information system -- Reference Model. OAIS reference model developed by the Consultative Committee for Space Data Systems (CCSDS)

ISO 16363

- ISO standard: Space Data for Information Transfer Systems – Audit and Certification of Trustworthy Digital Repositories
- Developed by the Consultative Committee for Space Data Systems (CCSDS) Mission Operations and Information Management Services (MOIMS) group
  - ISO Technical Committee (TC) 20 / Subcommittee (SC) 13
- Based on the requirements, Trusted Repositories Audit and Certification: Criteria and Checklist (TRAC), developed by the National Archives and Records Administration (NARA) and the Research Libraries Group (RLG)
- ISO Status: PRF
  - Proof of a new international standard

ISO 16919

- ISO standard: Space Data for Information Transfer Systems – Requirements for bodies providing Audit and Certification of Candidate Trustworthy Digital Repositories
- Developed by the Consultative Committee for Space Data Systems (CCSDS) Mission Operations and Information Management Services (MOIMS) group
  - ISO Technical Committee (TC) 20 / Subcommittee (SC) 13
- Based on ISO 16363, Space Data for Information Transfer Systems – Audit and Certification of Trustworthy Digital Repositories
- ISO Status: DIS
  - Approved by the CCSDS. ISO Draft International Standard (DIS)

ISO 16363 Test Audits

- Completed to test the use of ISO 16363 for the audit process, to identify improvements for ISO 16363 and ISO 16919, and to attain experience conducting audits
- Conducted in June and July 2011
- Conducted by the Trustworthy Digital Repositories Accreditation Board (PTAB), a subset of the CCSDS MOIMS group that developed the draft ISO 16363 standard
- Six volunteer archives audited
  - 3 in Europe (located in the UK, France, and Netherlands)
  - 3 in the United States (located in New York, Maryland, and Kentucky)

SEDAC Motivated to Participate in ISO 16363

- **Extrinsic Motivation**
  - Improving international standards for scientific data stewardship
  - Contributing to integrity and transparency of science
  - Enabling capabilities to ensure future use of scientific data
- **Intrinsic Motivation**
  - Identifying opportunities for continuous improvement of strategies, policies, plans, processes, and practices
  - Earning trust and obtaining independent recognition as compliant with international standard
  - Obtaining external assurances for communities of data providers, collaborators, users, and sponsors

SEDAC Plan for ISO 16363 Compliance

- SEDAC is planning to meet the requirements for ISO 16363 compliance by initially serving as a volunteer repository for the test audits that are being conducted to examine the application of the draft standard for use in the audit and certification of trustworthy digital repositories.
- Based on the results of the test audit conducted in mid 2011 and on recommendations received from the volunteer team of test auditors, SEDAC identified improvements needed to meet the expected requirements for ISO 16363 compliance.

SEDAC Preparation for Test Audit

- Enhancements resulting from previous external audits and reviews
- Self-assessment of SEDAC Long-Term Archive, recommended by LTA Board, conducted in 2009, presented at Open Repositories 2009 and published:
  - Downs and Chen (2010) Self-Assessment of a Long-Term Archive for Interdisciplinary Scientific Data as a Trustworthy Digital Repository. *Journal of Digital Information*, 11(1), http://journals.tdl.org/jodi/article/view/753
- Improvements identified for procedures and practices
- Scope of SEDAC ISO 16363 test audit, SEDAC characteristics, and preparation for audit presented during PTAB Meeting at NARA on February 18, 2011
- Self audit spreadsheet reporting SEDAC compliance with ISO 16363 submitted to PTAB on June 16, 2011

PTAB Test Audit Site Visit of SEDAC

- Site visit conducted on June 20 – 21, 2011
- Site visit team consisted of 7 volunteer PTAB test auditors from the US and Europe
  - R. Downs recused himself from PTAB site visit team for audit of SEDAC
- Brief overview of SEDAC presented by R. Chen
- Questions on various aspects of metrics compliance
- Review of mission, policies, plans, procedures, records, and documentation
- Inspections of facilities, competencies, and capabilities
- Observations of operations and practices
- Verbal report received from PTAB at end of second day

Issues of Privacy and Non-Disclosure

- Some information and documents describing security plans and details about systems, infrastructure, operations and finance are confidential and cannot be duplicated or distributed outside of SEDAC.
- The names of individual SEDAC staff members or their titles will not be included in any reports that are published or distributed beyond the team that conducts the audit.
- Partnering organizations, collaborators, and their staff members must not be identified in any reports that are published or distributed beyond the team that conducts the audit.
- Investigators obtained approval from Columbia University's Morningside campus Institutional Review Board (IRB) to study audit of SEDAC*

* R. Downs is vice-chair of this IRB, but was recused from this decision

SEDAC Areas Audited

- **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

Initial Results of ISO 16363 Test Audit of SEDAC

- **General observations**
  - Strong tradition of adding value to data evident
  - Good attention to succession planning for preservation
  - Good representation on User Working Group (UWG), including maintaining domain of expertise of UWG membership
  - UWG review of dissemination data prior to dissemination
  - Commitment to archiving input data sets for value added products
  - High level of staff interactions with Producers and Consumers by participation in various user and standard groups
  - Good use of standards such as Geospatial Standards

Initial Results of ISO 16363 Test Audit of SEDAC

- **Recommendations for improvement**
  - Define the long-term “Designated Community” for each collection
  - Develop clearer distinctions between the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP), especially when implementing Fedora
  - Ingest scientists should capture history of actions as they convert SIPs
  - Information associated with each AIP needs to be enhanced, especially for future migration efforts
  - Representation Information should be captured down to a file level
  - Capture fixity information early during acquisition process and increase integrity checking through the data lifecycle
  - Improve systematic approach to preservation planning including technology watch and monitoring of formats, software, etc. in the archive
  - Physically separate storage of archival masters from circulation copies of data to reduce risk of mishandling
  - Improve training of scientific and other staff in archival procedures

Next Steps

- Review recommendations received from test audit
- Establish plan for addressing recommendations for improvement
- Implement improvements according to plan
- Apply for ISO 16363 Certification after implementing improvements