Independent Evaluation of a Scientific Data Center for Compliance with the ISO 16363 Requirements for Audit and Certification of Trustworthy Digital Repositories

Robert R. Downs [1] and Robert S. Chen[2]
Center for International Earth Science Information Network (CIESIN),
Columbia University,

[1]rdowns@ciesin.columbia.edu
[2] bchen@ciesin.columbia.edu

Prepared for Presentation to the
Research Data Alliance (RDA) Second Plenary Meeting
September 16, 2013,
National Academy of Sciences,
Washington, DC

Objectives: Scientific data centers and other digital repositories need to continuously improve so that they can meet the challenge of providing stewardship for the scientific data that are used by scientists, policy-makers, educators and their students, and the general public. As part of its efforts to continuously improve its capabilities and services offered to communities that are interested in using scientific data on human interactions in the environment, SEDAC, the NASA Socioeconomic Data and Applications Center, requested an independent test audit to evaluate its compliance with the draft requirements of ISO 16363:2012, the international standard for Audit and Certification of Trustworthy Digital Repositories. SEDAC has conducted various audits through the years as part of its self-improvement efforts. However, obtaining an independent test audit for compliance with the draft metrics in ISO 16363 offers an opportunity to identify ways in which the scientific data center could improve its organization and management, its processes for managing and disseminating data, and its systems and security infrastructure. Similarly, when organizations and auditors are authorized to offer certification of trustworthy digital repositories, SEDAC will be better prepared to apply for certification as a result of having been previously audited by independent evaluators for compliance with the draft ISO 16363 metrics.

On-going activities: SEDAC initially prepared for the ISO 16363 test audit by conducting various internal evaluation activities for continuous quality improvement (CQI), including an internal audit of the SEDAC Long-Term Archive (LTA) for compliance with the requirements in the TRAC document, Trustworthy Repositories Audit & Certification: Criteria and Checklist. Completing these ongoing evaluation and improvement activities, including the internal TRAC audit of the LTA, identified various opportunities for improving policies, plans, procedures, and documentation, which were revised and adopted as part of the CQI efforts. The scope of the ISO 16363 test audit was determined to include the entire management and operations of SEDAC. Limitations were identified for access to certain financial and security documents by the external auditors. The scope and limitations were presented to the group that developed ISO 16363, along with information about SEDAC and its policies, in preparation for the SEDAC site visit. The self-assessment questionnaire, describing how SEDAC addressed the ISO 16363 requirements, also was completed and submitted to the auditors prior to the site visit. Completing the self-assessment questionnaire also provided SEDAC with an opportunity to further improve its processes to ensure that the draft requirements of ISO 16363 were addressed for all SEDAC management and operations.

ISO 16363 Metrics Where Improvement is Needed	Corresponding Areas Being Reviewed and Improved
3.1 Governance and Organizational Viability	Mission statement and policies - emphasize commitment to continuing stewardship and preservation of scientific data and services. Plans for transferring data, operations, responsibilities, and authority to another entity in case of an unforeseen event Preservation plans to include details of new procedures as they are adopted
3.2 Organizational Structure and Staffing	Data stewardship training to be completed by new staff and periodically by experienced staff, which includes Open Archival Information Systems (OAIS) standards and terms
3.3 Procedural Accountability and Preservation Policy Framework	Processes to define designated community for each AIP during data development and data dissemination planning
4.1 Ingest: Acquisition of Content	Procedures for recording all inventory, verification, and maintenance activities performed on objects and collections
4.2 Ingest: Creation of the Archival Information Package	Procedures for testing and improving the understandability of each AIP for the designated community Procedures for recording the provenance of activities completed during data development and dissemination
4.3 Preservation Planning	Procedures to identify, record, and maintain information on software dependencies for each file received
4.4 Archival Information Package Preservation	Procedures to verify the integrity of digital objects and files
5.1 Technical Infrastructure Risk Management	Risk management plans to include an organizational risk register containing tracked risk mitigation schedules Procedures to separate circulation copies of AIPs from archival copies

Results: Seven external auditors conducted a two-day site visit of SEDAC as part of the independent test audit for compliance with the draft ISO 16363 requirements. The site visit included introductions, a description of the audit, a briefing on SEDAC, inspections of documents and facilities, observations of operations, interviews of staff members, verification of records of activities for consistency with policies and procedures, and a debriefing. Recommendations offered by the auditors included defining the designated community for each collection; differentiating between processes for the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP); capturing provenance when converting SIPs; enhancing information for each AIP; capturing fixity earlier; capturing representation information at the file level; improving preservation planning; separating archival masters from circulation copies; and improving training. The recommendations have led to the creation of an improvement plan and the development of enhancements, including adoption of the BagIt specification and DROID, data review and workflow modifications, and an assessment of the user community. By embracing opportunities to evaluate their practices on a continuing basis, scientific data centers can improve their data stewardship capabilities to meet the scientific data needs of current and future communities and become trustworthy digital repositories.

References

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

Downs, RR, Chen, RS. 2012. Improving the Trustworthiness of an Interdisciplinary Scientific Data Archive. IASSIST 2012.

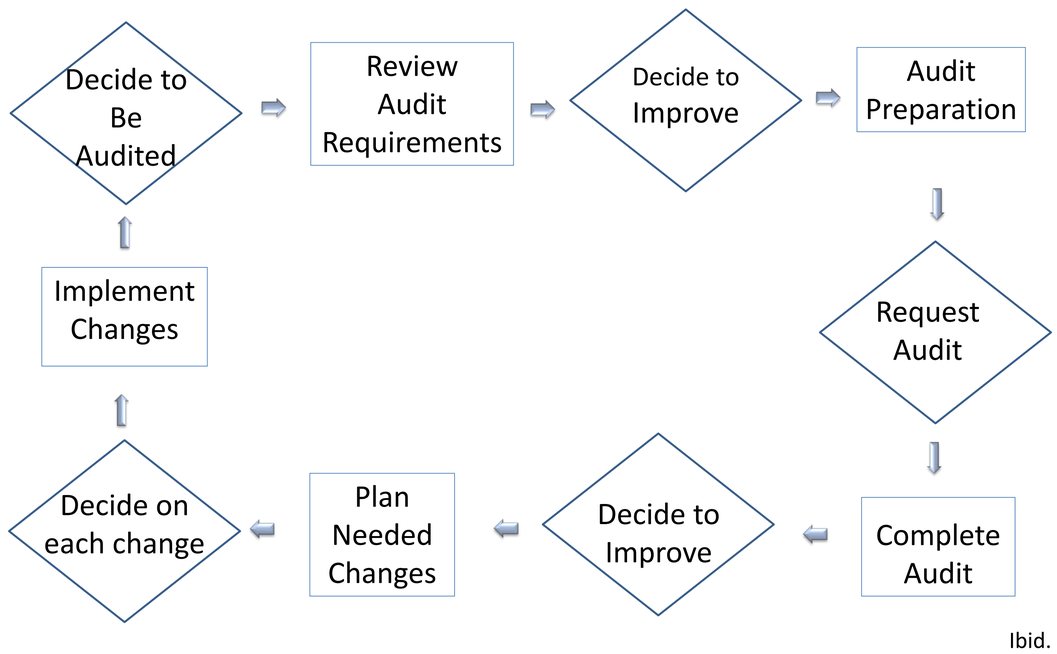
Downs, RR, Chen, RS. 2011. Audit of a Scientific Data Center for Certification as a Trustworthy Digital Repository: A Case Study. http://earthdata.nasa.gov/sites/default/files/webform/agu/DownsChenAuditTrustworthyCaseStudy20111209.pdf

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

The authors appreciate the support received from the National Aeronautics and Space Administration under Contract NNG08HZ11C for the Continued Operation of the Socioeconomic Data and Applications Center (SEDAC).

Downs & Chen, 2012

Continuously Improving the Scientific Data Archive



Ibid.

http://sedac.ciesin.columbia.edu/

The study was approved by the Columbia University Institutional Review Board as protocol AAAI1568.