

Assessing a Scientific Data Center as a Trustworthy Digital Repository

Robert R. Downs¹ and Robert S. Chen²

¹ rdowns@ciesin.columbia.edu

² bchen@ciesin.columbia.edu

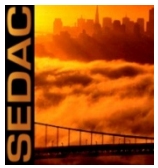
NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network
(CIESIN)
The Earth Institute, Columbia University

PREPARED FOR PRESENTATION TO THE
ESIP DATA STEWARDSHIP WORKING GROUP

Monday 20 April 2015, 2:00 p.m. ET



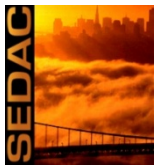
Using ISO 16363 to Assess a Trustworthy Digital Repository



- ISO 16363:2012
 - Published as the international standard for Audit and Certification of Trustworthy Digital Repositories
- Developed by the Consultative Committee for Space Data Systems (CCSDS)
 - Mission Operations and Information Management Services (MOIMS) area focusing on Repository Audit and Certification (RAC)
 - <http://public.ccsds.org/publications/archive/652x0m1.pdf>



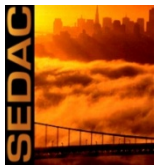
History & Current Status of ISO 16363 Certification



- Recommended practices developed
 - Consultative Committee for Space Data Systems (CCSDS) developed recommended practices based on:
 - Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC).
- International standards published
 - Space data and information transfer systems -- Audit and certification of trustworthy digital repositories. Published as ISO 16363:2012
 - Space data and information transfer systems -- Requirements for bodies providing audit and certification of candidate trustworthy digital repositories. Published as ISO 16919:2014
- Need to establish certifying bodies for each country
 - Currently, various organizations provide consulting, training, and audit services based on ISO 16363
 - No organizations have been accredited to provide ISO 16363 certification.



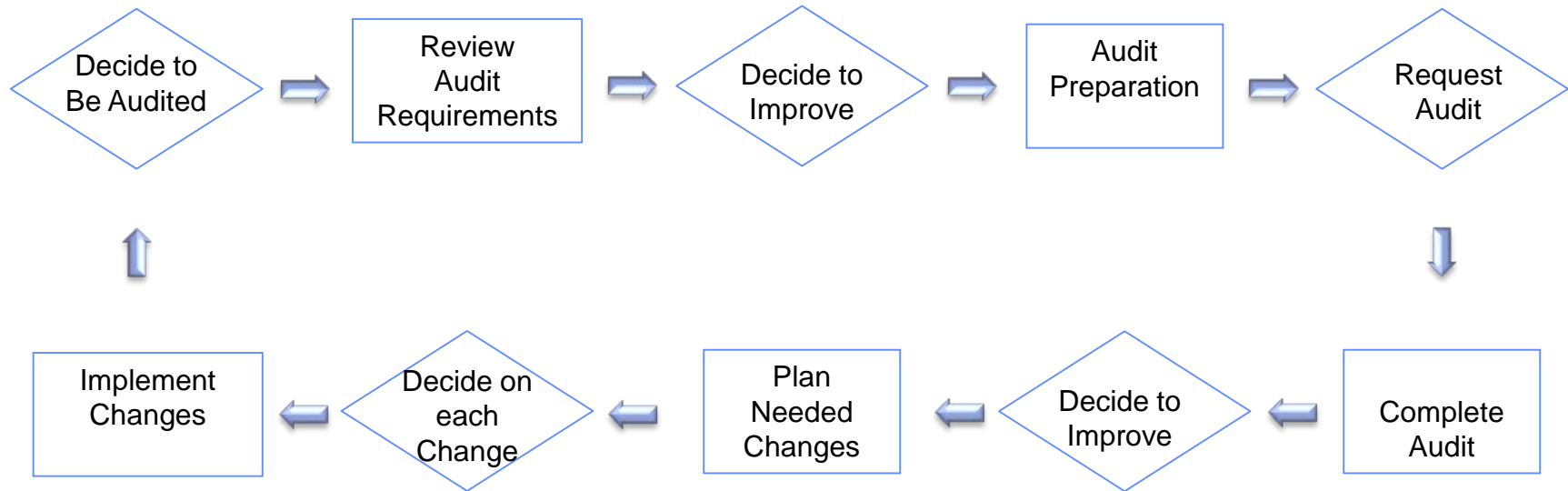
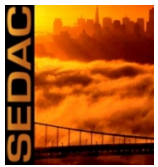
Conduct a Self-Assessment of ISO 16363 Compliance



- Familiarize staff with ISO 16363 Self-Assessment Template
 - Review each section: Organizational Infrastructure, Digital Object Management, Infrastructure and Security Risk Management
- Assign top-level metrics to staff with relevant responsibilities
 - Review each low-level metric to identify current practices and evidence
 - Revise policies and procedures to reflect current practices
 - Populate Evidence column with titles of documents and records
 - Provide explanation for how evidence addresses each metric
- Identify improvements needed to meet each metric
 - Review Evidence and Explanation columns and identify weaknesses
- Implement improvements to address each weakness
 - Develop, and adopt needed policies, procedures, tools, and techniques
 - Strengthen staff capabilities via professional development and hiring
- Revise each section for consistency with adopted improvements
 - Revise Evidence and Explanation to reflect documents and practices

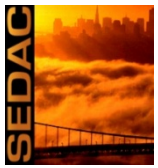


Continuously Improving the Scientific Data Archive





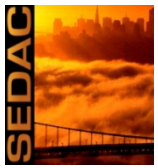
ISO 16363 Test Audits



- 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 Primary Trustworthy Digital Repositories Accreditation Board (PTAB), a subset of the CCSDS MOIMS RAC 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)
 - Represent a range of natural and social science and humanities archives and data centers



NASA Socioeconomic Data and Applications Center (SEDAC) at Columbia University



- SEDAC is one of twelve Distributed Active Archive Centers (DAACs) in the NASA Earth Observing System Data and Information System and one of the original ESIP Federation Type 1 members (1999)
- SEDAC focuses on interdisciplinary datasets useful to the research, applications, and education user communities
- SEDAC has been operated by CIESIN, a research center of the Earth Institute at Columbia University, since 1998 through a series of 5-year contracts
- CIESIN has an agreement with the Columbia University Libraries for long-term stewardship for SEDAC data
- ISO 16363 certification has been of particular interest since SEDAC data tend to fall outside of traditional agency, library, and science community preservation responsibilities
- ISO 16363 is also of interest in light of SEDAC application to become part of the ICSU World Data System (WDS), replacing former CIESIN status as a World Data Center)

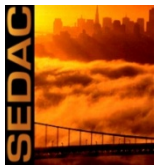
The screenshot shows the SEDAC website interface. At the top, there is a navigation menu with links for DATA, MAPS, THEMES, RESOURCES, COMMUNITY, ABOUT, and HELP. Below the menu is a search bar and a navigation bar with the text "NASA SOCIOECONOMIC DATA AND APPLICATIONS CENTER (SEDAC) - HOSTED BY CIESIN AT COLUMBIA UNIVERSITY". The main content area features a map titled "Map Layer: Human Footprint Version 2" showing a global map with color-coded regions. To the right of the map is a sidebar with the text "Last of the Wild, Version 2" and a list of links: "Data Collection", "Data Sets (2)", and "Map Gallery". Below the map, there are sections for "Featured Uses of Data", "More Featured Uses...", "News", and "More News...". The "Featured Data Sets" section lists "Population Density Grid, v1 (1990, 1995, 2000)", "Global Rural-Urban Mapping Project (GRUMP), v1", "Dams, v1.01 (2011)", and "Global Reservoir and Dam (GRAD), v1". Each data set has an "Overview" and "Download" link. The "Recent Releases" section lists "Global Footprint and Mosaic", "Global Rural-Urban Mapping Project (GRUMP)", and "Terrestrial/SEDAC Viewer". The footer contains the SEDAC logo, a NASA logo, and the text "Copyright © 1997-2012. Privacy".

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





ISO 16363 Test Audit of SEDAC



- **Prior to site visit**

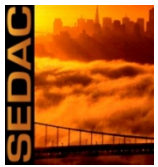
- Presented to Primary Trustworthy Digital Repository Authorization Board (PTAB) on scope of SEDAC audit and on preparation activities completed
- Provided PTAB with a report on self-assessment of SEDAC compliance with ISO 163636 requirements
- PTAB reviewed self-assessment and online policies, procedures, and reports

- **Site visit of SEDAC**

- Conducted for 2 consecutive days by 7 members of PTAB from US and Europe
- Managers and staff were informed of audit process and interviewed in a group
- Facilities inspected and printed documents requested and reviewed
- Staff observed and questioned while conducting routine operational activities
- Performance of specific tasks requested and observed with detailed questions
- Mutual debriefing by site visit team and SEDAC managers at end



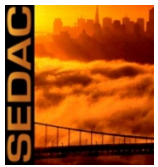
SEDAC Areas Audited



- **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



Observations from ISO 16363 Test Audit of SEDAC

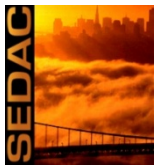


- **General Observations**

- 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



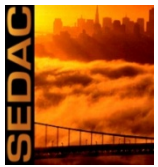
ISO 16363 Metrics Requiring Improvement



- 3.1 Governance and Organizational Viability
- 3.2 Organizational Structure and Staffing
- 3.3 Procedural Accountability and Preservation Policy Framework
- 4.1 Ingest: Acquisition of Content
- 4.2 Ingest: Creation of the Archival Information Package
- 4.4 Archival Information Package Preservation
- 5.1 Technical Infrastructure Risk Management



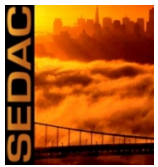
Areas Identified for Improvement



1. Mission statement and policies - **emphasize commitment to continuing stewardship** and preservation of scientific data and services
2. **Plans for transferring data, operations, responsibilities, and authority** to another entity in case of an unforeseen event
3. Preservation plans to **include details of new procedures** as they are adopted
4. Data stewardship **training to be completed by new staff and periodically by experienced staff**, which includes Open Archival Information Systems (OAIS) standards and terms
5. Processes to define designated community for each AIP during data development and data dissemination planning
6. Procedures for recording all inventory, verification, and maintenance activities performed on objects and collections
7. Procedures for **testing and improving the understandability** of each AIP for the designated community
8. Procedures for **recording the provenance of activities** completed during data development and dissemination
9. Procedures to **identify, record, and maintain information on software dependencies** for each file received
10. Procedures to **verify the integrity** of digital objects and files



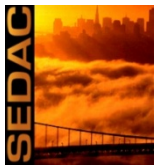
SEDAC Improvement Plan for ISO 16363 Compliance (1 of 2)



1. Review and improve mission statement and policies to emphasize the commitment to continuing stewardship and preservation of scientific data and services.
2. Review and improve plans for transferring data, operations, responsibilities, and authority to another entity in case of an unforeseen event.
3. Review and improve data stewardship training to be completed by new staff and periodically by experienced staff, which includes Open Archival Information Systems (OAIS) standards and terms.
4. Review and improve processes to define the designated community for each Archival Information Package (AIP) as part of data development and data dissemination planning.
5. Review and improve procedures for testing and improving the understandability of each AIP for the designated community.



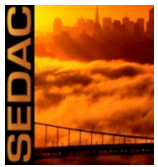
SEDAC Improvement Plan for ISO 16363 Compliance (2 of 2)



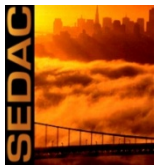
6. Review and improve procedures for recording the provenance of activities completed during data development and dissemination processes.
7. Review and improve procedures to identify, record, and maintain information on software dependencies for each file received.
8. Review and improve procedures for recording all inventory, verification, and maintenance activities performed on objects and collections.
9. Review and improve procedures to separate circulation copies of AIPs from archival copies.
10. Review and improve preservation plans to include details of new procedures as they are adopted.
11. Review and improve risk management plans to include an organizational risk register containing tracked risk mitigation schedules.



Progress on Implementing Improvements



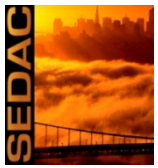
- Adopted the **BagIt Specification** for packaging content and preservation information into the AIP and replaced proprietary integrity checking software with the open source BagIt Library created by the Library of Congress.
- Deployed the open source **DROID tool** and established procedures to record and maintain information on software dependencies for every file contained in each AIP
- Revised the **data dissemination process** to generate each Dissemination Information Package (DIP) from the AIP
- Created **training modules** that can be used for developing data stewardship capabilities of data managers, data developers, and systems engineers
- Revised data **appraisal, assessment, and review processes** to ensure that evaluation, planning, and quality control activities are complete and documented during data development and dissemination processes
- Implemented **Digital Object Identifiers (DOIs)** to persistently identify each disseminated dataset and improve the citation of each disseminated dataset
- Conducting an **assessment of the SEDAC user community**, including the identification of disciplines served and current trends



- Repositories that serve **interdisciplinary and diverse** communities face substantial challenges to meet the requirements of ISO 16363
- Increasing the **preparation for an audit** improves the quality of recommendations received for further improvement
- Assessment and improvement are an **ongoing process** that may require incremental enhancements to attain success and to maintain the momentum for achieving progress
- A **system perspective** of the archive is needed to understand the relationships among the requirements for trustworthy repositories



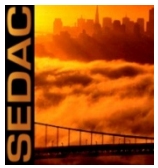
Insight Gained from Test Audit Process (2 of 2)



- Plans for improvement should identify opportunities for **increasing efficiency** while increasing operational effectiveness
- As understanding of what it means to be a trustworthy repository improves, the requirements for meeting each metric should increase, requiring **continuous improvement**
- As knowledge on trustworthy repositories increases, available technologies and capabilities to meet the requirements should increase, requiring **perpetual learning**
- Plans for improvement should attempt to **surpass standards**, which serve as a benchmark representing the minimum level of achievement that is necessary for scientific data stewardship



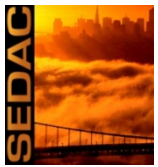
Benefits of ISO 16363 Audit of SEDAC



- Improves **transparency** of archival processes
- Improves **quality assurance** for stakeholders
- Provides **independent evaluation** of the archive by external archival experts
- Improves **efficiency and effectiveness** of archival operations
- Recommendations guide **planning** for archival enhancements
- Improves **management and stewardship of data** maintained by archive
- Increases **data preservation capabilities** of archive
- Measures **compliance with requirements** of international standard
- Recognition of **responsibilities and achievements** of data center staff
- Necessary step for **certification** as a trustworthy digital repository
- Supports SEDAC's application to join the **ICSU WDS**



References



- **Audit and Certification of Trustworthy Digital Repositories. Recommended Practice.** 2011 CCSDS <http://public.ccsds.org/publications/archive/652x0m1.pdf>
- **Audit Preparation and Self-Assessment Template for ISO 16363.** 2013 PTAB <http://www.iso16363.org/preparing-for-an-audit/>
- Downs RR, Chen RS 2013 **Independent Evaluation of a Scientific Data Center for Compliance with the ISO 16363 Requirements for Audit and Certification of Trustworthy Digital Repositories.** RDA 2nd Plenary, <http://hdl.handle.net/10022/AC:P:21793>
- Downs RR, Giaretta D, Hughes JS 2014 **Policy Recommendations for Institutions to Serve as Trustworthy Stewards of Research Data.** RDAP 2014, <http://dx.doi.org/10.7916/D8SB43SP>
- **Reference Model for an Open Archival Information System (OAIS).** 2012 CCSDS <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- **Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories.** 2014 CCSDS <http://public.ccsds.org/publications/archive/652x1m2.pdf>
- **Trustworthy Repositories Audit & Certification: Criteria and Checklist.** 2007 CRL http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf