

## **Interagency Science Working Group**

### **National Archives and Records Administration<sup>1</sup>**

#### **Establishing Trustworthy Digital Repositories: A Discussion Guide Based on the ISO**

#### **Open Archival Information System (OAIS) Standard Reference Model**

**January 19, 2011**

Federal agencies often require long-term storage and access to electronic records, digital data, and other information needed for business, but often have not put into place the fundamental policy and governance infrastructure to assure the trustworthiness and authenticity of the data they maintain for as long as it is needed. As required by the National Archives and Records Administration (NARA) in 36 CFR Chapter XII, Subchapter B, Records Management, Federal agencies are responsible for creating and maintaining authentic, reliable, and usable records and ensure that they remain so for the length of their authorized retention period. This guide, developed by NARA in collaboration with a team of archivists, records managers, data managers and scientists, was created to help Federal agencies determine what high-level data management policies, procedures, and processes are in place now, and what policies, processes and procedures need to be updated or developed to assure that long-term digital data is appropriately managed, preserved, and made accessible to all stakeholders for as long as it is needed.

## **BACKGROUND**

---

<sup>1</sup> NARA's Interagency Science Working Group is composed of records management, professional, scientific, and archives staff from NARA, U.S. Geological Survey, National Oceanic and Atmospheric Administration, Environmental Protection Agency, and National Aeronautics and Space Administration.

NARA's Interagency Science Working Group is comprised of experienced archivists, records managers, data managers, and scientists in the Federal Government. In FY 2010, the group was charged with developing guidance, tools, checklists, best practices, or similar resources for Federal science agencies that would help them acquire, store, manage, preserve, and provide access to scientific data, and especially large volumes of digital observational data, over long periods of time. The team decided that a high-level discussion guide for agency managers and staff about the Open Archival Information System (OAIS) standard reference model would be useful before an agency undertakes a long-term, large-scale project to implement the OAIS Standard, known as ISO 14721:2003. The group found that a higher level discussion guide also builds on work already going on in some scientific agencies to implement the ISO standard, and may help other Federal agencies with long-term storage, preservation, and access needs to develop policies, procedures, and practices that help preserve the records and assure their authenticity and trustworthiness throughout the entire lifecycle of the records and data. The group also built on a "best practice" the National Oceanic and Atmospheric Administration (NOAA) developed to comply with the OAIS standard and NARA records management regulations for the OAIS Ingest process. The *NOAA Procedure for Scientific Records Appraisal and Archive Approval: A Guide for Data Managers*, and an abbreviated brochure for program officials and managers, are available at <http://nosc.noaa.gov/docs/products.html>. This appraisal procedure implements NOAA Administrative Order 212-15, Management of Environmental and Geospatial Data and Information, NOAA's high-level guidance for procedures, decisions, and actions regarding environmental data and information management throughout the agency, available at [http://www.corporateservices.noaa.gov/~ames/NAOs/Chap\\_212/naos\\_212\\_15.html](http://www.corporateservices.noaa.gov/~ames/NAOs/Chap_212/naos_212_15.html).

## THE OAIS STANDARD

The OAIS standard is a widely recognized benchmark used by many national and international scientific, cultural and research organizations, institutions, and agencies to develop appropriate governance, policies, business processes, and procedures to operate a data archive for long-term preservation, storage and access to digital information. It is independent of any particular scientific or academic discipline or field of study, and is not restricted to an exclusively technical environment, as several of its components have a human or manual element. For the purposes of this discussion, a trustworthy digital repository is a repository whose mission is to acquire and preserve digital data so it is able to provide reliable, long-term access to the managed digital resources to its designated community (or communities), now and in the future. The term “data” is used throughout to identify all digital information, including that which meets the definition of a Federal record found in 44 U.S.C. 3301. The business processes or functions listed in this guide are the same “functional entities” referred to in the OAIS standard.

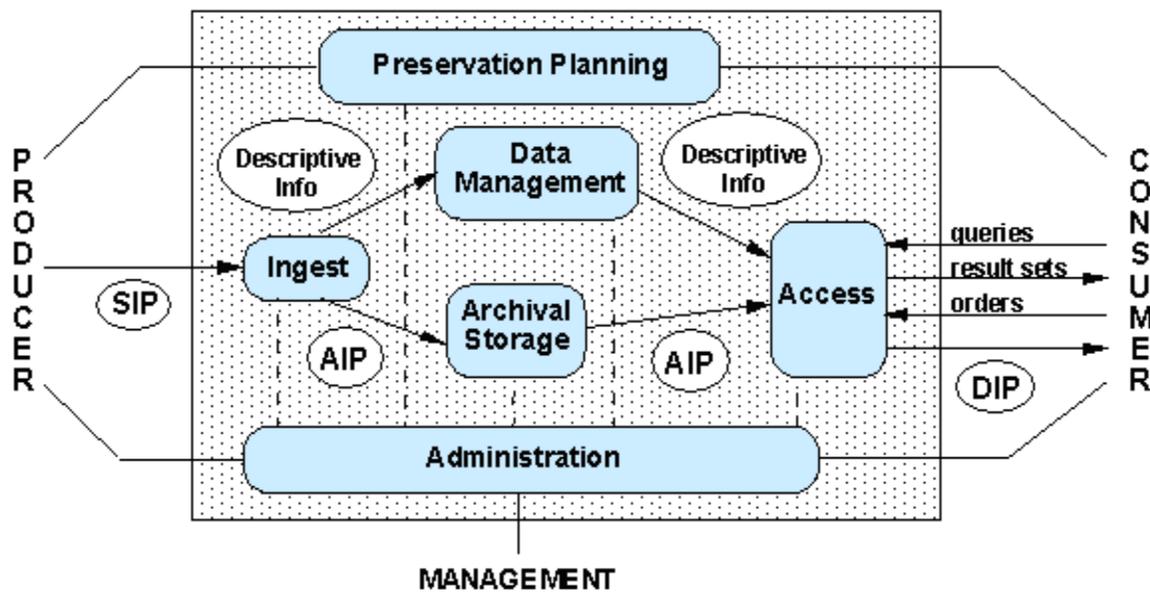
The OAIS standard reference model is based on six inter-related and overlapping “functional entities” which are referred to in this guide as “business processes.” They are:

- (1) Ingest,
- (2) Archival Storage,
- (3) Data Management,
- (4) Administration,
- (5) Preservation Planning, and
- (6) Access.

Three external clients or stakeholders interact with these processes: producers, consumers, and management (see Figure 1, below, and the reference sources for a glossary and additional detailed information). Overall, the standard reference model treats the operation of an

OAIS-compliant digital data archive or trusted digital repository as a management entity that requires dedicated governance (financial resources, personnel, policies, procedures, oversight, and compliance). The model focuses on the processes required for the long-term preservation and accessibility of digital objects in an archival framework but does not specify the nature of the implementation of such an archive system. The reference model is independent of any particular scientific or academic discipline or field of study, and is also not restricted to an exclusively technical environment, as several of its components have a human or manual element.

**Figure 1: OAIS Standard Reference Model, Interrelationships of OAIS Business Processes**



Source: Procedures Manual for the Consultative Committee for Space Data Systems (2001)

*Acronyms: SIP: Submission Information Package; AIP: Archival Information Package; DIP: Dissemination Information Package*

## HOW YOU CAN USE THIS GUIDE

This guide provides a series of high-level, open-ended questions related to each of the six OAIS business processes. The questions are designed to help a team or group of agency managers, staff, and subject matter experts identify policy, procedures, and operational areas for which agency guidance exists, where gaps in policies, procedures and operations exist, and where revisions to policies and procedures are needed to meet OAIS criteria for long-term preservation and access to digital records and information.

The questions should be discussed in a meeting or interview with agency staff to assure that everyone understands the processes and the underlying questions, and is familiar with OAIS terminology. To answer the questions, it may be helpful to select one dataset (or one continuing series of datasets that are the same) and walk through your data repository's lifecycle processes, from data creation, storage or maintenance, access, preservation, to eventual disposal or transfer to NARA. It may be necessary to consult with or conduct interviews with a team of subject matter experts in your agency, including the agency records manager, agency data managers, enterprise architects, program managers, scientists, and others chosen by your agency to help respond to the questions. Some information may be readily available and easy to obtain, and other information may require more in-depth interviews and research.

Using this guide can benefit your agency by improving communications among knowledgeable staff regarding specific issues, questions or processes related to long-term preservation and access to digital data. A wide variety of public websites, documents, policies, or other materials are readily available that may also inform your responses, provide additional information, or identify a "best practice" for a process that your agency can implement.

## DISCUSSION QUESTIONS

**Ingest Process:** This entity or business process provides the services and processes to accept Submission Information Packages (SIPs) from creators or producers and prepare the contents for storage and management within the archive. These functions and activities include, but are not limited to, identifying and receiving submissions, conducting quality assurance, generating the initial Archival Information Packages (AIPs; documentation and the object itself), generating initial descriptive information, and coordinating updates as needed.

1. Data acquisitions (or accessions) to a repository:
  - a. Does your agency have formal criteria or a collection and appraisal policy that identifies the types of data collected and maintained?
  - b. Is there a formal procedure to submit, review and appraise data submissions to determine if they warrant ingest into the archive?
  - c. Are the data listed on a record schedule or schedules approved by NARA (the agency records officer is a good contact for this information), and if so, what are the approved dispositions?
  - d. If the records are appraised by NARA as permanent, are they created and maintained in a format that meets NARA transfer standards for electronic records?
2. Does the repository have a formal, defined quality assurance program to verify ingest functions, validate the file(s), and maintain data file and metadata integrity?
3. Has your agency defined what metadata must be saved when the data are initially created? Does the repository have additional or supplementary metadata requirements?

4. Are future research uses and business needs evaluated in the initial review of the data?  
Could this data be used for research that is not related to its original purpose? For example, meteorological data are used to forecast the weather, the primary value of the data, but are often used later by attorneys and insurance companies to verify claims, or by engineers to design wind and adverse weather loads for new buildings, the secondary value of the data.
5. Does the repository obtain both legal title and physical custody of ingested datasets created outside your organization? How is this documented? Are restrictions on data disclosure (e.g. privacy, security classification, confidential business information, FOIA exemptions, etc.), identified in the data submission, prior to ingest?
6. What are the most significant current challenges in your agency's repository for the ingest function (e.g., financial resources, personnel, policies, procedures, oversight, and compliance)? To what extent do these challenges impact your agency's ability to conduct the activities in this process? How could NARA advocacy and guidance help your repository or agency address these challenges?

**Archival Storage Process:** This entity or business process provides the services and functions for the storage, maintenance and retrieval of Archival Information Packages (AIP) including the data objects. Functions and activities in this entity include, but are not limited to, receiving datasets and creating final approved AIPs, managing data storage and hardware and software, determining the impact of hardware and software upgrades, replacing media, conducting error checks, providing disaster recovery services, and providing data for access.

1. How does your repository evaluate metadata associated with the data? Has your repository developed metadata standards, or are other recognized international or widely

accepted metadata standards used to facilitate management, access, and retrieval? What is the source of the standard(s) your repository uses, e.g., ISO, ANSI, etc.

2. Is there a policy or formal procedure to assure that all media used to store data is reviewed regularly for replacement, data loss, and data failures?
3. Is there an approved comprehensive disaster recovery plan with current implementation guidance for the repository? Does it cover technical failures as well as natural disasters and other failures that are identified in the security and risk management plans for the repository?
4. What are the most significant current challenges in your agency's repository for the archival storage function (e.g., financial resources, personnel, policies, procedures, oversight, and compliance)? To what extent do these challenges impact your agency's ability to conduct the activities in this process? How could NARA advocacy and guidance help your repository or agency address these challenges?

**Data Management Process:** This entity or business process provides the services and functions for populating, maintaining, and accessing indexes and catalogs that provide descriptive information that identifies and documents archive holdings as well as databases of internal archive administrative data. Typical functions and activities in this area include, but are not limited to, administering databases, performing queries, generating reports, receiving database updates, and the like.

1. Does your repository have a procedure to assure that the index, catalog, or database is updated if descriptive or administrative information is changed? Does your repository have a procedure to track new versions of datasets that are submitted by data creators? Is there a policy that covers what the repository does with older versions of data files that have been migrated to newer formats, or requires upgraded or different software or hardware to access?
2. Are there policies in place governing who has access to create, read, use and delete the descriptive information and administrative data?
3. Are the databases searchable, or can searches be done across a variety of datasets? How accessible is the data your agency manages to internal staff and external customers? Does your index or catalogue require special expertise to use and understand?
4. What are the most significant current challenges in your agency's repository for the data management function (e.g., financial resources, personnel, policies, procedures, oversight, and compliance)? To what extent do these challenges impact your agency's ability to conduct the activities in this process? How could NARA advocacy and guidance help your repository or agency address these challenges?

**Administration Process:** This entity or business process manages the overall operation of the archive system (e.g., all six OAIS business processes), including establishing standards, developing archive policies and procedures, negotiating with clients, identifying new and continuing data submissions, managing system configuration, updating archival information for long-term preservation, maintaining physical security, auditing SIPs and AIPs for conformity to archive policies and procedures, managing risk, providing customer services, and the like.

1. Are there formal procedures and policies for ingest, data appraisal, data preservation, and for developing long-term preservation approaches and strategies that include risk management, disaster recovery, security, audits and quality checks, data transformation or data migration activities? If so, are they documented and available to all users with a need for them?
2. Does your repository comply with any widely recognized standards for creating and maintaining data and related metadata such as Dublin Core, ISO standards, Federal Geographic Data Committee (FGDC) standards, Federal Information Processing Standards (FIPS), etc?
3. Are there submission agreements in place with data producers? If so, are they documented and widely available? Does your repository maintain an ingest schedule for continuing datasets, or rely on the data creators (producers) to initiate an SIP?
4. What organizational entity has oversight of the repository for program accountability and operations? Are related organizational entities included or aligned with the overall accountability operations?
5. What, if any, functions or processes are performed by outside contractors? What are the contractual obligations relating to data management?
6. Are there internal audit policies in place to evaluate repository operations against goals or performance measures? If so, how often and to whom are the results provided to?
7. Is there a panel or team identified to review data submissions and/or archive holdings to assure they continue to meet business and customer needs?
8. What are the most significant current challenges in your agency's repository for the administration function (e.g., financial resources, personnel, policies, procedures,

oversight, and compliance)? To what extent do these challenges impact your agency's ability to conduct the activities in this process? How could NARA advocacy and guidance help your repository or agency address these challenges?

**Preservation Planning Process:** This entity or business process monitors the environment of the OAIS repository and provides guidance and recommendations to ensure that the information stored in the OAIS repository remains accessible to the data user communities over the long-term even if the original computing and data environment becomes obsolete.

1. Are recognized data preservation standards used, or are there formal agency or organizational policies and procedures in place to document the physical environment in the facility and the status of the computing environment? Are there formal policies and procedures to assure that all of the data in the archive is monitored regularly for obsolescence, version control, and updates or additions from the creators?
2. Does your agency regularly review customer and data user needs and requirements? Are emerging technologies regularly monitored to help identify potential preservation and access issues or improve data archive services and operations?
3. What are the most significant current challenges in your agency's repository for the preservation planning function (e.g., financial resources, personnel, policies, procedures, oversight, and compliance)? To what extent do these challenges impact your agency's ability to conduct the activities in this process? How could NARA advocacy and guidance help your repository or agency address these challenges?

**Access:** This entity or business process supports customers and internal or external stakeholders in searching, locating and requesting data descriptions and archival information products stored in the OAIS repository.

1. Is there a formal data dissemination and access plan that determines who has access to what data in the repository?
2. Are there regular efforts made to assess how well the repository serves its stakeholders, and to identify new and changing needs of different user groups, stakeholders, or consumers over time? Does your repository have user reporting policies and procedures to monitor access operations?
3. How do users access the archive's data index or catalog? How is the data searched and made available to consumers and other stakeholders?
4. How does your repository protect data from unlawful disclosure or protect privacy and other legal rights, and still provide access to the data (e.g., does your repository maintain "public use" versions?) Are there audit policies to assure that the information is appropriately protected from unauthorized disclosure? Are encryption techniques used to protect data in storage, or to deliver data to customers? Is the encryption scheme fully documented?
5. How does the repository determine what format the data is made available in? Does the repository use NARA's guidance on acceptable transfer formats, or are other generally accepted preservation and access formats used? Can the data be disseminated in open formats, or is proprietary software required on the part of the user? What open formats are used in your repository?

6. What are the most significant current challenges in your agency's repository for the ingest function (e.g., financial resources, personnel, policies, procedures, oversight, and compliance)? To what extent do these challenges impact your agency's ability to conduct the activities in this process? How could NARA advocacy and guidance help your repository or agency address these challenges?

### **Reference Resources and Links**

The latest "Pink Book" draft revision of the OAIS standard is available at <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/CCSDSAgency.aspx>. Please consult this document for OAIS terminology used in this survey.

The official ISO standard for OAIS, ISO 14721:2003, is available for purchase at [http://www.iso.org/iso/iso\\_catalogue.htm](http://www.iso.org/iso/iso_catalogue.htm).

The latest "Red Book" Recommended Practice for Audit and Certification of Trusted Digital Repositories, October, 2010, which is based on the OAIS standard, is available at <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206521R1/CCSDSAgency.aspx>. This is one of the documents this guide is largely based upon.

A 2007 report leading to the audit and certification of trusted repositories draft cited above. Center for Research Libraries, Online Computer Library Center, and the National Archives and Records Administration. Trustworthy Repositories Audit and Certification: Criteria and Checklist, February, 2007. Available at <http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-digital-repositories>.

*NOAA Procedure for Scientific Records Appraisal and Archive Approval: A Guide for Data Managers*, and an abbreviated brochure for program officials and managers, are available at <http://nosc.noaa.gov/docs/products.html>.

OAIS-related Website Links of Interest:

<http://nost.gsfc.nasa.gov/isoas/>

<http://lockss.stanford.edu/lockss/OAIS>

<http://www.icpsr.umich.edu/dpm/dpm-eng/foundation/tdr/compliance.html>

[http://en.wikipedia.org/wiki/Open\\_Archival\\_Information\\_System](http://en.wikipedia.org/wiki/Open_Archival_Information_System)

<http://easydigitalpreservation.wordpress.com/2009/07/29/oais-reference-model-part-i-background-and-influence/>

<http://www.clir.org/pubs/reports/pub116/sec4.html>

<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/using-oais-curation>