OAIS Reference Model Standard: Motivation, Applicability, Follow-on Efforts

NIST Conference on Long Term Knowledge Retention (LTKR): Archival and Representation Standards

Don Sawyer /NASA/NSSDC

March 15, 2006
Outline of Talk

- Brief OAIS Overview
- Motivation and Applicability
- Follow-on Efforts
Purpose and Scope

- Framework for understanding and applying concepts needed for long-term digital information preservation
  - Long-term is long enough to be concerned about changing technologies
  - Starting point for model addressing non-digital information
- Provides set of minimal responsibilities to distinguish an OAIS from other uses of ‘archive’
- Framework for comparing architectures and operations of existing and future archives
- Basis for development of additional related standards
- Addresses a full range of archival functions
Applicability

- Applicable to:
  - All long-term archives and
  - Organizations and individuals dealing with information that may need long-term preservation

- Does NOT specify an implementation
Document Organization

- Introduction
  - Purpose and Scope, Applicability, Rationale, Road Map for Future Work, Document Structure, and Definitions of Terms

- OAIS Concepts and Responsibilities
  - High level view of OAIS functionality and information models
  - OAIS external environment
  - Minimum responsibilities to become an “OAIS”

- Detailed Models
  - Functional model descriptions and information model perspectives

- Preservation perspectives
  - Media migration, compression, format conversions, and access service preservation

- Archive Interoperability
  - Criteria to distinguish types of cooperation among archives

- Annexes
  - Scenarios of existing archives, compatibility with other standards
- Information is always expressed (i.e., represented) by some type of data
- Data interpreted using its Representation Information yields Information
- Information Object preservation requires clear identification and understanding of the Data Object and its associated Representation Information

![Diagram showing the relationship between Data Object, Representation Information, and Information Object.](attachment:/diagram.png)
OAIS Archival Information Package

- **Package Description**: e.g., Information supporting customer searches for AIP

- **Archival Information Package (AIP)**:
  - **Content Information**: e.g., • Hardcopy document
  - • Document as an electronic file together with its format description
  - • Scientific data set consisting of image file, text file, and format descriptions file describing the other files

- **Packaging Information**: e.g., How to find Content information and PDI on some medium

- **Preservation Description Information (PDI)**: e.g., • How the Content Information came into being, who has held it, how it relates to other information, and how its integrity is assured
OAIS Functional Entities

SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package
Reference Model Status

- Widely adopted as starting point in digital preservation efforts
  - Digital libraries (e.g., Netherlands National Library)
  - Traditional archives (e.g., US National Archives)
  - Scientific data centers (e.g., National Space Science Data Center)
  - Commercial Organizations (e.g., Aerospace Industries Association preservation working team)
- Published as final CCSDS and ISO (14721:2003) standards
- CCSDS version is available at:
How did Effort Start?

ISO TC20, SC 14 was looking for additional standards to develop
- Participant from NASA/Jet Propulsion Laboratory (JPL) proposed development of standard archive format
- Was aware of extensive format standards developed by the Planetary Data System managed by JPL

Assignment was transferred to ISO TC 20, SC 13 in 1995
- Address data used in conjunction with space missions
- Address intermediate and indefinite long term storage of digital data

Consultative Committee for Space Data Systems
- International group of space agencies
- Developed variety of science discipline-independent standards
- Became working body for an ISO TC 20/ SC 13 about 1990

TC20: Aircraft and Space Vehicles
SC13: Space Data and Information Transfer Systems
Panel 2 (Information Interchange) Analysis

- Archives in different science domains, even within a single Space agency like NASA, used different terminology and had different format standards.

- Adoption of a single archive format would not be accepted among NASA archives, let alone internationally.

- Emergence of Web was enabling many data producers to make their data directly available.
  - Who needs an archive?
  - My floppy disk is an archive!

- Several Panel 2 participants recognized the growing digital preservation problem.

- What might be done?
Panel 2 responded to CCSDS and ISO TC20/SC 13

- No framework widely recognized for developing specific digital archive standards
- Therefore begin by developing a ‘Reference Model’ to establish common terms and concepts
- Ensure broad participation, including traditional archives

(Not restricted to space communities; all participation is welcome!)

- Focus on data in electronic forms, but recognize that other forms exist in most archives
- Follow up with additional archive standards efforts as appropriate
Organizational Approach

- Organized US contribution under a framework with NASA lead in 1995
  - Established liaison with Federal Geographic Data Committee (FGDC) and National Archives and Records Administration (NARA)
  - Agency archives and users must be represented in this process

- An “Open” process
  - Important to stimulate dialogue with broad archive/user communities
  - Results of US and International workshops put on WEB
  - Support e-mail comments/critiques

- Broad international workshops also held
  - UK and France
  - Issue resolution at ISO/Consultative Committee for Space Data Systems international workshops
Results

- Reference Model targeted to several categories of reader
  - Archive designers
  - Archive users
  - Archive managers, to clarify digital preservation issues and assist in securing appropriate resources
  - Standards developers

- Adopted terminology that crosses various disciplines
  - Traditional archivists
  - Scientific data centers
  - Digital libraries

- Become extremely well known as starting point for discussions addressing digital preservation
Follow-on Activities

- OAIS Reference Model has stimulated follow-on activities within CCSDS and outside

- CCSDS/ISO Producer-Archive Interface Methodology Standard (ISO 20652)
  - Provides framework for Producer/Archive interactions
  - Identifies steps and types of information exchanged during the ‘negotiation’
  - May be used as a checklist by archives

- CCSDS/ISO Producer-Archive Interface Specification
  - Draft standard addressing formal modeling and instantiation of Submission Information Packages

- RLG/NARA Digital Repository Certification Task Force
  - Produced a draft audit checklist for Repository Certification