Proceedings of the 9th Annual
Federal Depository Library Conference

October 22 - 25, 2000

The Open Archival Information System: A Model for Preserving Digital Information

Don Sawyer, NASA/NSSDC
Lou Reich, CSC
Greenbelt, MD

Outline

- What is a reference model?
- Why an OAIS reference model?
- Development of the OAIS reference model
- Highlights of the model
- Results to date
- Recommendations

What Is a Reference Model?

- A framework

  - for understanding significant relationships among the entities of an environment, and
  - for the development of consistent standards or specifications supporting that environment

- A reference model can also be called an abstraction of the key concepts, and their relationships, of an environment
- A reference model

  - is based on a small number of unifying concepts and
  - may be used as a basis for education and explaining standards to a non-specialist

Why an OAIS Reference Model?

- Agencies have a significant stewardship responsibility for the information obtained from their observational programs

  - Observational data are often irreplaceable
  - Taxpayer's investment must be prudently managed

- Long term (indefinite) preservation of this information is difficult

  - Data+metadata (i.e., information) must be migrated across new media, operating systems, and management systems
  - Field representations and formats may need to be revised to keep pace with evolving technologies and supported standards
  - What constitutes adequate metadata is not widely understood or standardized
  - Information is becoming ever more widely distributed
  - Information must be readily transportable from archive to archive

More on 'Why an OAIS Reference Model'

- No consensus on what 'archiving' means, or what services might be available from an archive

  - Preserving the information is not the same as preserving the data bits
  - Will the customer understand the information in 10, 50 or 100 years?
  - How can we know when effective archiving is being achieved?
  - Lack of consensus impedes establishment of standards and commercial support services

- ISO TC20 requested that an archive data standard be developed

  - Would be wasted effort without an agreed framework among users and archives
  - First need an archive reference model

Response to ISO

- No framework widely recognized for developing specific digital archive standards

- Establish an Archive Work-Package within CCSDS Panel 2 (Information Interchange)

- 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

US Organizational Approach

- Organize US contribution under a framework with NASA lead

  - Establish liaison with Federal Geographic Data Committee (FGDC) and NARA
  - Agency archives and users must be represented in this process
  - US contributions to be submitted to CCSDS Panel 2

- Will be 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

- Expect there will be a core group willing to devote 10-20% of time

  - Develop material and attend meetings

Getting Started

- First open US workshop held October 1995

  - Advertised widely
  -Variety of government, academic, and industry participation, including National Archives
    - Received presentations on archive efforts underway
    - Laid out initial reference model objectives, proposed way to work
    - Provided initial concept paper on what the reference model might look like
    - Confirmed that a reference model would be useful

  - Active US working group was formed

    - About 15-20 persons from a variety of agencies/organizations

  US Workgroup Takes Lead

  - US workgroup activities are fully open

    - Held 2-day working meetings approximately quarterly
    - New participants always welcome
    - Plans, minutes, drafts available from Web

  - Broad international workshops also held

- Britain and France

- Issue resolution at CCSDS international workshops

Technical Approach

- Investigated other Reference Models

  - ISO "Seven Layer"Communications Reference Model
  - ISO Reference Model for Open Distributed Processing
  - ISO TC211 Reference Model for Geomantics

- Define what is meant by ‘archiving of observational data’
- Break ‘archiving’ into a few functional areas (e.g., for ingest, storage, dissemination, and searching functions)
- Define a set of interfaces between the functional areas
- Define a set of data classes for use in archiving
- Choose formal specification techniques

  - Data flow diagrams for functional models and interfaces
  - Object Modeling Technique (OMT) for data classes

Resulting Model

- 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
  - Getting favorable comments wherever exposed

Reference Model for an Open Archival Information System

Open Archival Information System (OAIS)

- Open

  - Reference Model standard(s) are developed using a public process and are freely available

- Information

  - Any type of knowledge that can be exchanged
  - Independent of the forms (i.e., physical or digital) used to represent the information
  - Data are the representation forms of information

- Archival Information System

  - Hardware, software, and people who are responsible for the acquisition, preservation and dissemination of the information
  - Additional OAIS responsibilities are identified later and are more fully defined in the Reference Model document

Document Organization

- Introduction

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

- OAIS Concepts

  - 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

- Migration perspectives

  - Media migration, compression, and format conversions

- Archive Interoperability

  - Criteria to distinguish types of cooperation among archives

- Annexes

  - Scenarios of existing archives, compatibility with other standards

Purpose, Scope, and Applicability

- 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
- Applicable to all long-term archives and those organizations and individuals dealing with information that may need long-term preservation
- Does NOT specify any implementation

Model View of an OAIS’s Environment

- Producer is the role played by those persons, or client systems, who provide the information to be preserved
- Management is the role played by those who set overall OAIS policy as one component in a broader policy domain

- Consumer is the role played by those persons, or client systems, who interact with OAIS services to find and acquire preserved information of interest

OAIS Information Definition

- Information is defined as any type of knowledge that can be exchanged, and this information is always expressed (i.e., represented) by some type of data
- In general, it can be said that "Data interpreted using its Representation Information yields Information"
- In order for this Information Object to be successfully preserved, it is critical for an archive to clearly identify and understand the Data Object and its associated Representation Information

Information Package Definition

- An Information Package is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI)

OAIS Responsibilities

- Negotiates and accepts Information Packages from information producers
- Obtains sufficient control to ensure long-term preservation
- Determines which communities (designated) need to be able to understand the preserved information
- Ensures the information to be preserved is independently understandable to the Designated Communities

- Follows documented policies and procedures which ensure the information is preserved against all reasonable contingencies
- Makes the preserved information available to the Designated Communities in forms understandable to those communities

Reference Model Summary

- Reference model is to be applicable to all digital archives, and their Producers and Consumers
- Identifies a minimum set of responsibilities for an archive to claim it is an OAIS
- Establishes common terms and concepts for comparing implementations, but does not specify an implementation
- Provides detailed models of both archival functions and archival information
- Discusses OAIS information migration and interoperability among OAISs

Reference Model Status

- CCSDS Reference Model Red Book released August 1999

  - http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

- ISO Draft International Standard (DIS) released June 2000

  - Same content as CCSDS Red Book

- Comments are actively solicited

  - Participate in various ways:

    - ISO TC20/SC 13 review in your country
    - CCSDS review by your space agency

  - All comments will be considered and non-editorial comments will get a response

- Some comments received suggest we may want to add a process model addressing preservation

Results to Date

- Conferences to publicize/enhance the Reference Model
- Workshop Presentations to discuss Long Term Preservation and additional standards
- Partnership with Traditional Archives and Digital Libraries

  - Research Efforts
  - Use as Architecture

- Enhanced Communications among varied Communities

Conferences and Workshops

- Digital Archive Directions (DADs) workshop June 1998

  - http://ssdoo.gsfc.nasa.gov/nost/isoas/dads/

- Tutorial session at GSFC/IEEE Mass Storage Conference March 1999
- Archival Workshop on Ingest, Identification, and Certification (AWIICS) October 1999

  - http://ssdoo.gsfc.nasa.gov/nost/isoas/awiics/ws.html

    - Example presentations at various Conferences

      - Society of American Archivists 1997 annual meeting
      - NAGARA (Gov't archivists) 1998 Annual Meeting
      - NSF Workshop on Data Archival and Information Preservation, March 1999
      - ICSTI May meeting, presented by Gail Hodge

    Partnership with Traditional Archives and Digital Libraries

    - NARA and NASA were primary contributors to the OAIS RM and sponsors of DADS and AWIICS
    - Long article by Brian Lavoie in Online Computer Library Center (OCLC) summarizing the OAIS Reference Model
    - CEDARS: A multi-site UK project to create exemplars in Digital Archiving is using OAIS representation data as the basis for research into long term preservation
    - NEDLIB (Networked European Deposit Library) effort used OAIS RM as a basis for the design and architecture of Deposit System for Electronic Publications (DSEP)

    Enhanced Communications among Varied Communities

    - NARA contracted some work on long term preservation of collections to the San Diego Super Computer Center. Both parties claimed use of the OAIS RM saved about two weeks of effort in the specification of the task
    - Similar experiences between:

      - NCSA HDF format developers and DNA researchers
      - Life Sciences Archive developer and micro-gravity researchers
      - French space agency (CNES) and National Library of France representatives

Recommendations

- Look to the OAIS reference model to aid in managing and preserving digital resources

  http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf

- Critique the OAIS reference model by following instructions at:

  http://www.ccsds.org/RP9905/