Winter January 19, 2016

Session Two: OAIS Model & Digital Curation Lifecycle Model

Dr. Eun G Park
OAIS Model

- Reference Model for an Open Archival System in designing RMS
- Providing a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term.
Usage of OAIS Model

- 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

Purpose

- Standardize the relationships and interactions between an information Producer and an Archive Abstract Model
  - Need for formal standards for long-term storage of digital data
  - Need for consensus on requirements for maintaining digital content over the long-term, including: shared concepts/terminologies, basic functions, object attributes
Adoption of OAIS Model

- widely adopted as starting point in digital preservation efforts (Digital libraries, National Archives, Scientific data centers (e.g., National Space Science Data Center), Commercial Organizations (e.g., Aerospace Industries Association preservation working team)
- Lavoie, B. 2nd edition in October 2014

Model View of an OAIS Environment

**Producer** is the role played by 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 client systems who interact with OAIS services to find and acquire preserved information of interest
**OAIS Information Definition**

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

**Information Object**

- **Data Object** interpreted using its Representation Information yields **Information Object**.
- Information Object can be a Physical Object or a Digital Object, both of which are represented by Bit Sequence.
Representation Information

- The Representation Information accompanying a physical object may give additional meaning.
  - It typically is a result of some analysis of the physically observable attributes of the rock.

- The Representation Information accompanying a digital object, or sequence of bits, is used to provide additional meaning.
  - It typically maps the bits into commonly recognized data types such as character, integer, and real and into groups of these data types.
  - It associates these with higher level meanings which can have complex inter-relationships that are also described.

Recursive Nature of Representation Information

- Structure Information
- Semantic Information
- Other Representation Information

Structure Information * adds meaning to Representation Information
1
Interpreted using

Representation Information
1

Semantic Information
Other Representation Information

*
**Types of Information Used in OAIS**

- **Information Object**
  - **Content Information**
  - **Preservation Description Information**
  - **Packaging Information**
  - **Descriptive Information**

**Information Package Definition**

- An Information Package is a conceptual container holding two types of information:
  - Content Information
  - Preservation Description Information (PDI)
3 Information Package Variants

Submission Information Package (SIP)
- Negotiated between Producer and OAIS
- Sent to OAIS by a Producer

Archival Information Package (AIP)
- Information Package used for preservation
- Includes complete set of Preservation Description Information for the Content Information

Dissemination Information Package (DIP)
- Includes part or all of one or more Archival Information Packages
- Sent to a Consumer by the OAIS

SIP

- The information package that is transferred from the Producer to the OAIS when information is ingested into the archive.
- E.g.
  - The digital object and as much metadata as the Producer is willing or able to supply.
  - not be preserved in the exact form submitted by the Producer.
  - the aggregation of content provided in multiple SIPs
  - in a format not supported by the OAIS or necessitating migration to another format prior to inclusion in the archival store.
  - incomplete or inadequate form submitted by the producer may be augmented during the ingest process.
AIP

- The version of the information package that is stored and preserved by the OAIS.
- The focus of preservation, accompanied by a complete set of metadata sufficient to support the OAIS’s preservation and access services.
- E.g.
  - A single logical package within the archival system
  - Any form of physical association be maintained (embedding the metadata in the information object itself and storing the combined object as a single bit stream)

DIP

- The version of the information package delivered to the Consumer in response to an access request.
- The information package disseminated by the OAIS to the Consumer may differ in form or content to that which resides in the archival store
- E.g.
  - An image file might be converted from TIFF to JPEG prior to dissemination.
  - The amount of content may vary.
  - The DIP will not contain the complete set of metadata associated with an archived digital object, since much of it is of little interest to the Consumer.
External Data Flow View

Producer → Submission Information Packages

OAIS

Archival Information Packages

queries → result sets

Dissemination Information Packages

Consumer

OAIS: Six Functional Entities

Ingest → Archival Storage → AIP

Preservation Planning

Data Management

Descriptive Info.

Access

queries → result sets

orders → orders

Administration

SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package
Functional Entities in OAIS

- **Ingest**: to accept Submission Information Packages (SIPs) from Producers
- **Archival Storage**: to store, maintain, and retrieve Archival Information Packages (AIPs)
- **Data Management**: to maintain and access descriptive information which identifies and documents archive holdings and internal archive administrative data.
- **Administration**: to manage overall operation of the archive system
- **Preservation Planning**: to monitor and ensure that the information stored in the OAIS remain accessible to the Designated User Community over the long term
- **Access**: to request and receive information products

OAIS 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
OAIS Usage Examples

• Networked European Deposit Library (NEDLIB)
• Library of Congress
  – Hosting METS XML data packaging approach
  – National Digital Information Infrastructure Preservation Program (NDIIPP)
• InterPARES
  – Body of National Archives from many countries, adopted OAIS as a starting point for their modeling work
• France set up a working group within ARISTOTE
  – interested in archive of digital information, including libraries and Dept of Justice.
    • http://www.aristote.asso.fr/ (in french)
    • “astonishing unifying role” from OAIS reference model

The DCC (Digital Curation Centre) Curation Lifecycle Model

- A graphical **overview of the stages** required for curation and preservation of data from initial conceptualisation or receipt (2004).
- **Source:**
  

DCC Curation Lifecycle Model: Purposes

- To plan activities within an organization to ensure **all necessary stages are undertaken in the correct sequence**
- To define roles and responsibilities
- To build a **framework of standards and technologies to implement**
- To identify additional steps and actions in the process which may be required, or not required by certain situations or disciplines
- To ensure processes and policies are adequately documented.
The DDC Lifecycle Model consists of:

- Data – 2 kinds
- Full Lifecycle Actions - 4
- Sequential Actions - 8
- Occasional Actions - 3

Data (Digital objects or databases)

- Data, any information in binary digital form, is at the centre of the Curation Lifecycle. This includes:
- Digital objects as discrete form:
  - Simple Digital Objects - textual files, images or sound files, along with their related identifiers and metadata
  - Complex Digital Objects (by combining digital objects) - websites
- Databases: Structured collections of records or data stored in a computer system.
Full Lifecycle Actions 1

- **Description and Representation Information**: Assign administrative, descriptive, technical, structural and preservation metadata, using appropriate standards, to ensure adequate description and control over the long-term. Collect and assign representation information required to understand and render both the digital material and the associated metadata.

- **Preservation Planning**: Plan for preservation throughout the curation lifecycle of digital material. This would include plans for management and administration of all curation lifecycle actions.

Full Lifecycle Actions 2

- **Community Watch and Participation**: Maintain a watch on appropriate community activities, and participate in the development of shared standards, tools and suitable software.

- **Curate and Preserve**: Be aware of, and undertake management and administrative actions planned to promote curation and preservation throughout the curation lifecycle.
Sequential Actions 1

- **Conceptualise**: Conceive and plan the creation of data, including capture method and storage options.
- **Create or Receive**: Create data including administrative, descriptive, structural and technical metadata. Preservation metadata may also be added at the time of creation. Receive data, in accordance with documented collecting policies, from data creators, other archives, repositories or data centres, and if required assign appropriate metadata.
- **Appraise and Select**: Evaluate data and select for long-term curation and preservation. Adhere to documented guidance, policies or legal requirements.

Sequential Actions 2

- **Ingest**: Transfer data to an archive, repository, data centre or other custodian. Adhere to documented guidance, policies or legal requirements.
- **Preservation (Action)**: Undertake actions to ensure long-term preservation and retention of the authoritative nature of data. Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity. Actions include data cleaning, validation, assigning preservation metadata, assigning representation information and ensuring acceptable data structures or file formats.
Sequential Actions 3

• **Store**: Store the data in a secure manner adhering to relevant standards.

• **Access, Use and Reuse**: Ensure that data is accessible to both designated users and reusers, on a day-to-day basis. This may be in the form of publicly available published information. Robust access controls and authentication procedures may be applicable.

• **Transform**: Create new data from the original, for example
  - By migration into a different format.
  - By creating a subset, by selection or query, to create newly derived results, perhaps for publication.

Occasional Actions 1

• **Dispose**: Dispose of data, which has not been selected for long-term curation and preservation in accordance with documented policies, guidance or legal requirements. Typically data may be transferred to another archive, repository, data centre or other custodian. In some instances data is destroyed. The data’s nature may, for legal reasons, necessitate secure destruction.
Occasional Actions 2

- **Reappraise**: Return data which fails validation procedures for further appraisal and reselection.
- **Migrate**: Migrate data to a different format. This may be done to accord with the storage environment or to ensure the data’s immunity from hardware or software obsolescence.