The OAIS Reference Model: current implementations

Michael Day,
UKOLN, University of Bath
m.day@ukoln.ac.uk

Chinese-European Workshop on Digital Preservation,
Beijing, China, 14-16 July 2004

Presentation outline

- The OAIS Reference Model (brief)
  - Background
  - Definitions
  - Functional Model
  - Information Model
- Implementing the model:
  - Preservation metadata
  - System design
  - Conformance and certification

The OAIS reference model (1)

- Background:
  - *Reference Model for an Open Archival Information System (OAIS)*
  - Development led by the Consultative Committee for Space Data Systems (CCSDS)
  - Adopted as ISO 14721:2003
  - Establishes a common framework of terms and concepts
  - Identifies the basic functions of an OAIS
  - Defines an information model
  - www.ccsds.org/documents/650x0b1.pdf

The OAIS reference model (2)

- Provides definitions of terms that need to have well-defined meanings, e.g.:
  - Archival Storage, Content Data Object, Designated Community (key term), Ingest, Metadata, Representation Information, etc.
  - OAIS = "An archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community" (OAIS 1.7.2)

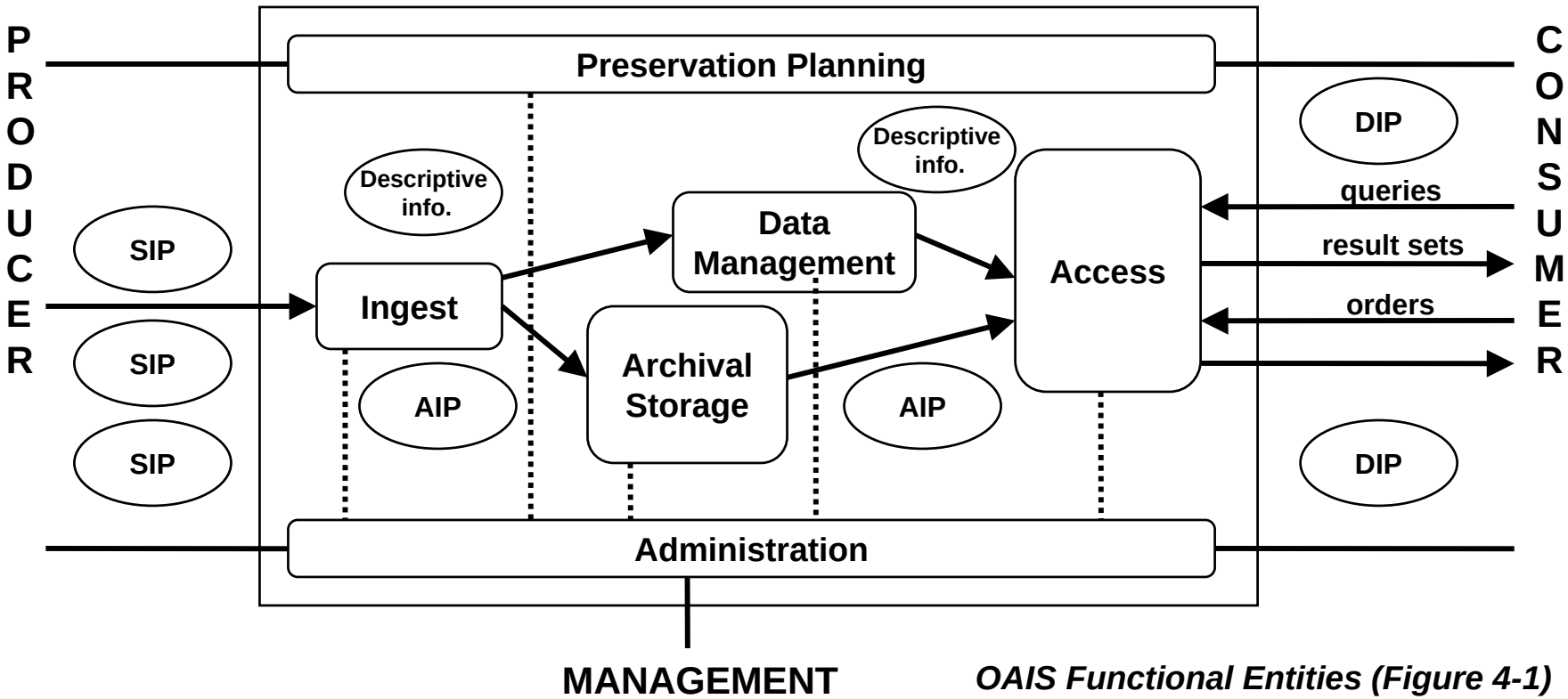
The OAIS reference model (3)

- High level concepts, e.g.:
  - The *environment* of an OAIS (Producers, Consumers, Management)
  - Definitions of *information*, Information Objects and their relationship with Data Objects
  - Definitions of *Information Packages*, conceptual containers of Content Information and Preservation Description Information
- OAIS Responsibilities

The OAIS reference model (4)

- Functional Model:
  - Six entities
    - Ingest
    - Archival Storage
    - Data Management
    - Administration
    - Preservation Planning
    - Access
  - Described using UML diagrams

The OAIS reference model (5)



OAIS Functional Entities (Figure 4-1)

The OAIS reference model (6)

- Information model:
  - Information Object (basic concept):
    - Data Object (bit-stream)
    - Representation Information (permits “the full interpretation of Data Object into meaningful information”)
  - Information Object Classes:
    - Content Information
    - Preservation Description Information (PDI)
    - Packaging Information
    - Descriptive Information

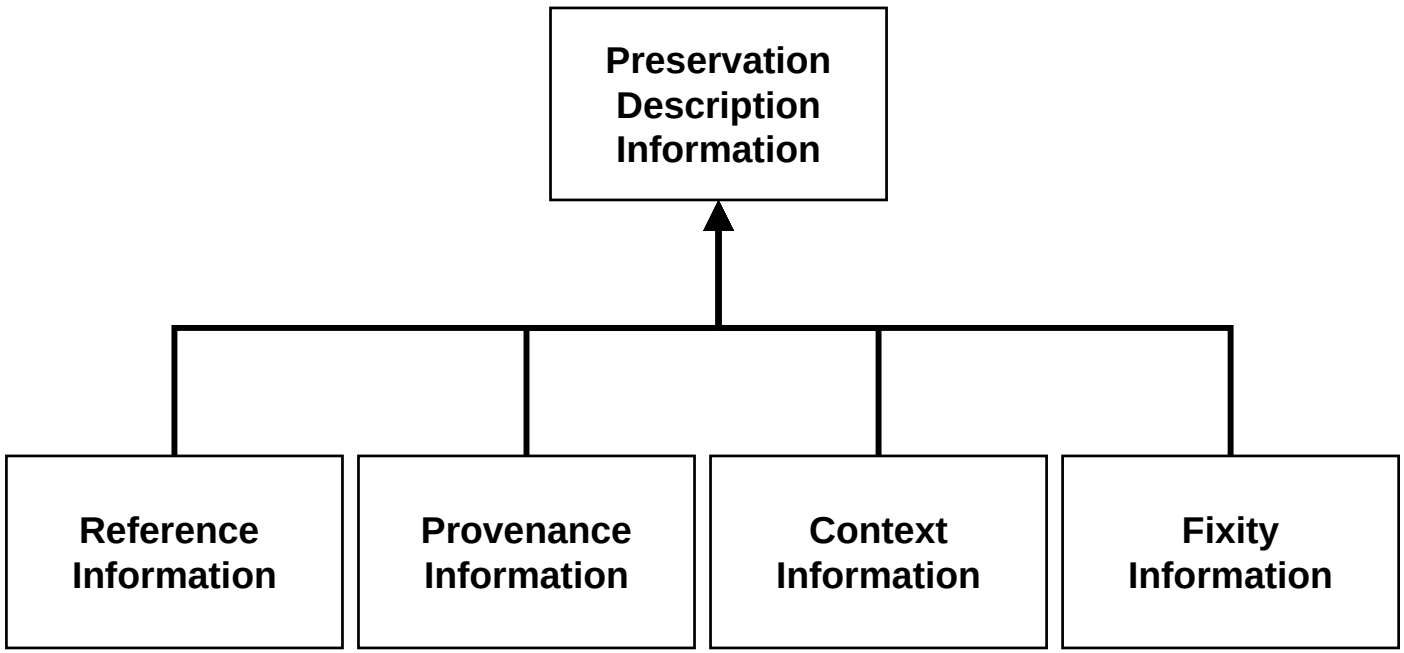
The OAIS reference model (7)

- Information model (continued):
  - Information package:
    - Container that encapsulates Content Information and PDI
    - Packages for submission (SIP), archival storage (AIP) and dissemination (DIP)
    - AIP = "... a concise way of referring to a set of information that has, in principle, all of the qualities needed for permanent, or indefinite, Long Term Preservation of a designated Information Object"

The OAIS reference model (8)

- Archival Information Package (AIP):
  - Content Information
    - Original target of preservation
    - Information Object (Data Object & Representation Information)
  - Preservation Description Information (PDI)
    - other information (metadata) “which will allow the understanding of the Content Information over an indefinite period of time”
    - A set of Information Objects
    - Based on categories discussed in CPA/RLG report: *Preserving Digital Information* (1996)

The OAIS reference model (9)



PDI Preservation Description Information (Figure 4-16)

The OAIS reference model (10)

- There is much, much more …
  - The Archival Information Unit and Archival Information Collections
  - Data transformations
  - Digital Migration
  - Preservation of look and feel (e.g., emulation, virtual machines)
  - Interoperability

Implementing OAIS (1)

- Fundamentals:
  - OAIS is a reference model (conceptual framework), NOT a blueprint for system design
  - It informs the design of system architectures, the development of systems and components
  - It provides common definitions of terms … a common language, means of making comparison
  - But it does NOT ensure consistency or interoperability between implementations

Implementing OAIS (2)

- ISO 14721:2003 published in early 2003
- However, earlier versions of the model made available by the CCSDS informed implementations long before then
- Three broad areas of influence:
  - Preservation metadata schemas
  - Architecture and system design
  - Conformance criteria

Implementing OAIS - metadata (1)

- The OAIS Information Model has been used to inform the development of preservation metadata schemas, e.g.:
  - Draft schemas developed by the National Library of Australia, Cedars project, NEDLIB project, etc.
  - Information Model used for the conceptual structure of the OCLC/RLG Metadata Framework (2002)
  - PREMIS working group

Implementing OAIS - metadata (2)

- The Metadata Encoding and Transmission Standard (METS)
  - An XML-based metadata framework for the association of digital object content with metadata
  - An XML 'wrapper' combining 'buckets' of descriptive, administrative and structural metadata
  - For object transfer or presentation
  - METS objects can be seen as an implementation of OAIS Information Packages (SIP, AIP, and DIP)

Implementing OAIS - systems (1)

- Two main uses (to date):
  1. To analyse existing preservation management practices
     - Helps with the comparison of repositories and the identification of important gaps
  2. "It is assumed that implementers will use this reference model as a guide while developing a specific implementation to provide identified services and content" (OAIS 1.4)
     - Some examples …

Implementing OAIS - systems (2)

- Ensuring Access to Mathematics Over Time
  - Cornell University Library and SUB Göttingen
  - Stated aim to "develop and implement a system that adheres to the principles put forth in the Open Archival Information System (OAIS) Reference Model."
  - Maintains directory of OAIS functions
  - www.library.cornell.edu/dlit/EATMOT/

Implementing OAIS - systems (3)

- Stanford Digital Repository
  - Stanford University Libraries & Academic Information Resources
  - Designing "OAIS-compliant" system for managing digitised objects, publications (e.g. from HighWire Press), electronic records, software, etc.
  - "… a means to gather and protect every facet of the University's organizational memory and intellectual capital."
  - Needs to scale to petabytes
  - www.diglib.org/pubs/news04_01/stanford.htm

Implementing OAIS - systems (4)

- OCLC Digital Archive Service
  - Subscription service provided by OCLC Online Computer Library Center, Inc.
  - Objects harvested from Web or can be submitted in batch (ingest)
  - Integrated with cataloguing workflows
  - OCLC preservation metadata specification
  - "Based on OAIS"
  - www.oclc.org/digitalarchive/

Implementing OAIS - systems (5)

- Many other examples:
  - Cedars project distributed archive prototype
    - Representation nets
  - Harvard University Library
    - XML-based Submission Information Package for e-journals
  - DSpace
  - KB *e*-Depot

Implementing OAIS - conformance (1)

- Many repositories claim OAIS compliance:
  - e.g., DSpace, OCLC Digital Archive, METS, etc.
  - Also, the LOCKSS System has produced a "formal statement of conformance to ISO 14721:2003" (lockss.stanford.edu/)
- The OAIS model claims to be a basis for conformance (OAIS 1.4), e.g.:
  - Supporting the information model (OAIS 2.2),
  - Fulfilling mandatory responsibilities (OAIS 3.1)

Implementing OAIS - conformance (2)

- OAIS Mandatory Responsibilities:
  - Negotiating and accepting information
  - Obtaining sufficient control of the information to ensure long-term preservation
  - Determining the "designated community"
  - Ensuring that information is "independently understandable"
  - Following documented policies and procedures
  - Making the preserved information available

Implementing OAIS - conformance (3)

- OCLC/RLG Digital Archive Attributes Working Group
  - Trusted Digital Repositories report (2002)
  - Recommended the development of a process for the certification of digital repositories
    - Audit model
    - Standards model
  - Goes well beyond OAIS mandatory responsibilities
    - e.g., administrative responsibility, organisational viability, financial sustainability, system security, etc.

Implementing OAIS - conformance (4)

- RLG/NARA Task Force on Digital Repository Certification
  - Research Libraries Group (RLG) and the US National Archives and Records Administration (NARA).
  - To define certification model and process
    - Identify those things that need to be certified (attributes, processes, functions, etc.)
    - Develop a certification process (organisational implications)
    - Draft proposal to be issued later in 2004

Summing up

- The OAIS model is a foundation stone for current and future digital preservation efforts
- It is already widely used to inform the development of preservation tools and repositories
- It could be used in the future as a basis for conformance, but its mandatory responsibilities do not identify all relevant criteria
- There may be a need for periodic revision, some clarification of definitions, possibly some comment from the archives world

Key links:

- OAIS Reference Model: http://www.ccsds.org/documents/650x0b1.pdf
- DPC Technology Watch Report on OAIS model by Brian Lavoie (OCLC Research): http://www.dpconline.org/
- RLG/NARA Task Force on Digital Repository Certification: http://www.rlg.org/

Acknowledgements

UKOLN is funded by Museums, Libraries and Archives Council, the Joint Information Systems Committee (JISC) of the UK higher and further education funding councils, as well as by project funding from the JISC, the European Union and other sources. UKOLN also receives support from the University of Bath, where it is based.

Also thanks to the Digital Preservation Coalition, the Digital Curation Centre, the DELOS Network of Excellence preservation cluster.