The OAIS Reference Model

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

Reference Models meeting, One Great George Street, London, 25 January 2006

Presentation outline

- The OAIS Reference Model
  - Background
  - Definitions, high level concepts, mandatory responsibilities
  - Functional Model
  - Information Model
- Implementing the model:
  - Preservation metadata
  - Informing system design
  - Conformance and certification

OAIS background

- Reference Model for an Open Archival Information System (OAIS)
  - Development led by the Consultative Committee for Space Data Systems (CCSDS)
  - Issued as CCSDS Recommendation (Blue Book) 650.0-B-1 (January 2002)
  - Also adopted as: ISO 14721:2003
  - http://public.ccsds.org/publications/archive/650x0b1.pdf

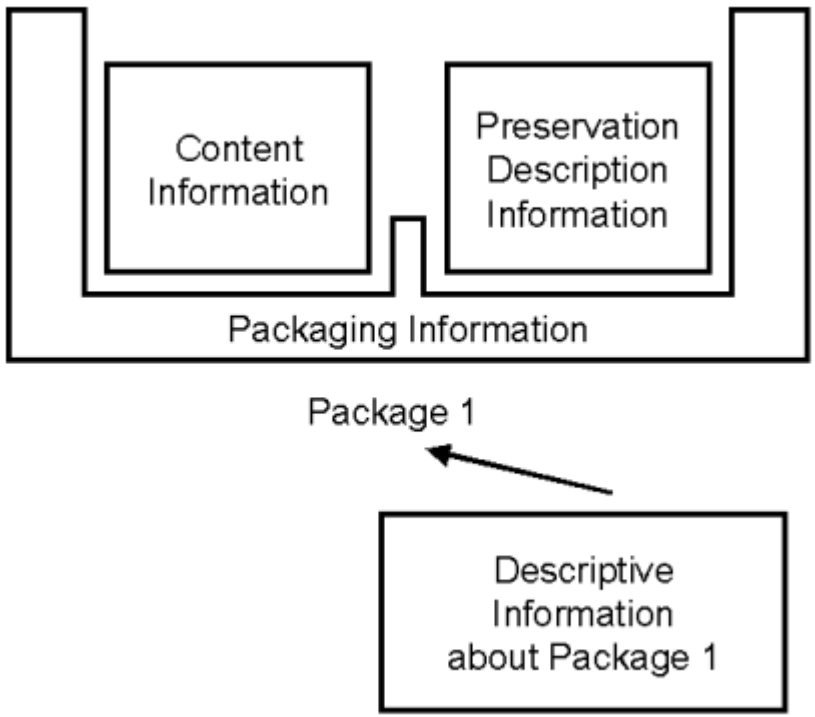
OAIS definitions

- 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)

OAIS high level concepts (1)

- 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 high level concepts (2)



Information Package Concepts and Relationships (Figure 2-3)

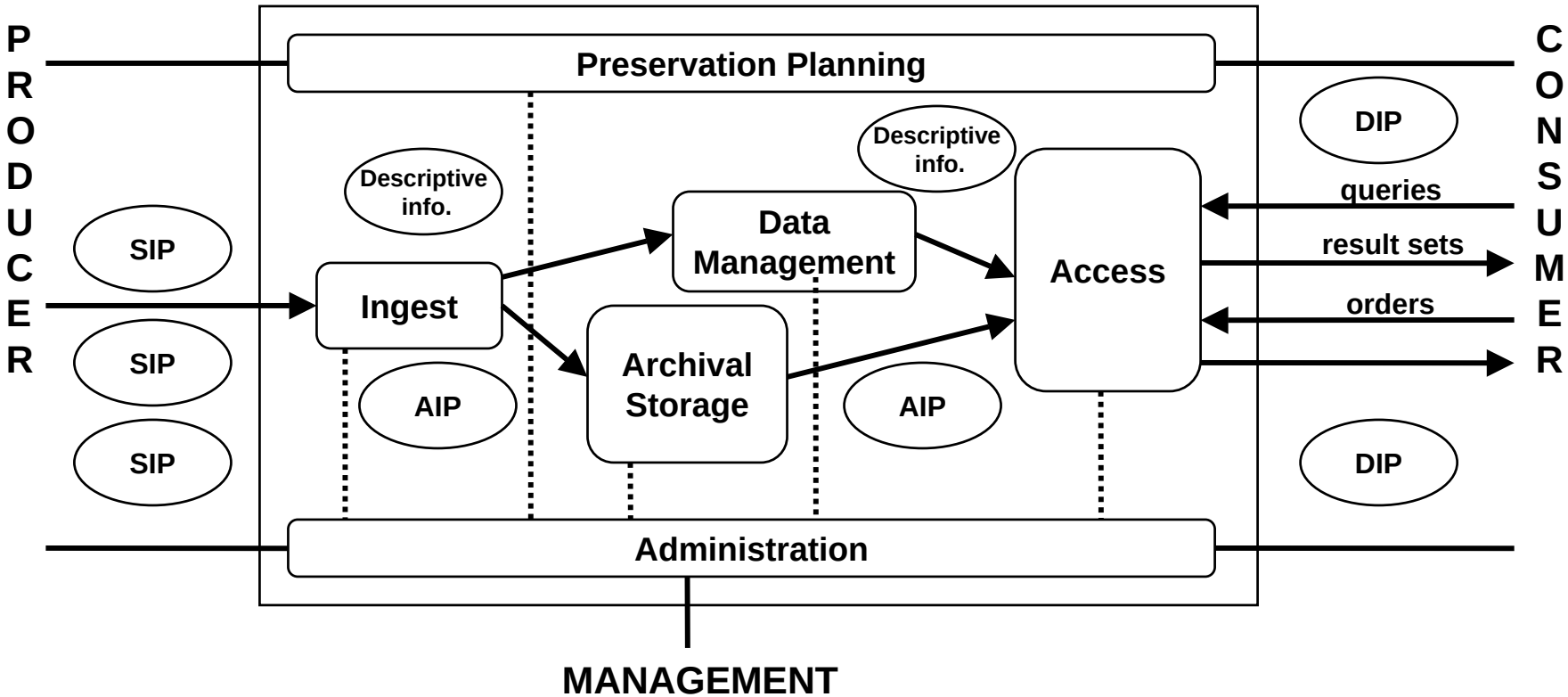
OAIS mandatory responsibilities

- 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**, i.e. without the assistance of those who produced it
- Following documented policies and procedures
- Making the preserved information available

OAIS Functional Model (1)

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

OAIS Functional Model (2)



OAIS Functional Entities (Figure 4-1)

OAIS Information Model (1)

- 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

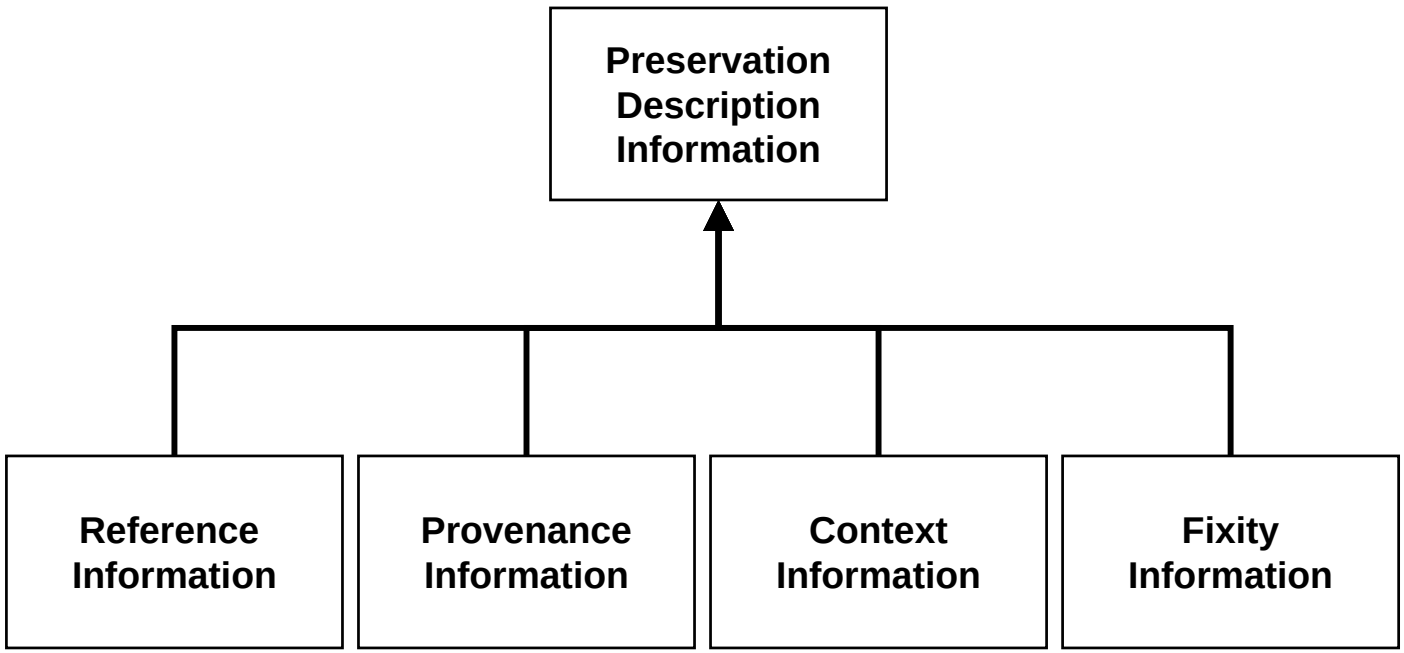
OAIS Information Model (2)

- 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"

OAIS Information Model (3)

- 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)

OAIS Information Model (4)



PDI Preservation Description Information (Figure 4-16)

OAIS Information Model (5)

- Also defines:
  - Archival Information Units and Archival Information Collections
  - Information Package transformations, e.g. for Ingest and Access
  - Preservation perspectives:
    - Migration, e.g refreshment, replication, repackaging, transformation
    - Preservation of look and feel (e.g., emulation, virtual machines)
  - Archive interoperability, e.g. federation

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 - follows the text made available by the CCSDS
- However, the 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 for repositories

Implementing OAIS - metadata

- The OAIS Information Model has been used to inform the development of many preservation metadata schemas, e.g.:
  - Draft schemas developed by the National Library of Australia, Cedars project, NEDLIB project, etc.
  - METS (Metadata Encoding and Transmission Standard) interpreted as an implementation of the OAIS Information Package concept
  - Information Model explicitly used for the structure of the OCLC/RLG Metadata Framework (2002)
  - Different approach taken by PREMIS Data Dictionary (2005)

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
     - Studies of BADC, UK Data Archive, …
  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)
     1. Examples …

Implementing OAIS - systems (2)

- Stanford Digital Repository
  - "OAIS-compliant" system for managing digitised objects
- OCLC Digital Archive Service
  - Subscription service claimed to be "Based on OAIS"
- Harvard University Library
  - XML-based Submission Information Package for e-journals
- Cedars project
  - Distributed archive prototype - Representation nets
- DCC RI Registry/Repository, DSpace, KB *e*-Depot, …

Implementing OAIS - conformance (1)

- Many repositories or preservation tools claim OAIS compliance:
  - e.g., DSpace, OCLC Digital Archive, METS
  - 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 checklist for self certification (August 2005), being tested by various projects in US, also by DCC

Some personal comments (1)

- Conformance with the OAIS model is often cited by digital preservation efforts (e.g. DSpace, METS) - but, given the nature of the model, can these claims be meaningful?
- At present, the model is best seen as a means of comparison between repositories, or a means of judging progress (e.g. Data Archive, BADC)
- Sometimes interpreters of the model seem reluctant to acknowledge that it could be improved

Some personal comments (2)

- OAIS mandatory responsibilities do not seem to identify all relevant criteria - but, together with the additional requirements now developed by the RLG-NARA Task Force, could be used in the future as a starting point for conformance
- There may be a need for periodic revision, some clarification of definitions, possibly some comment from the archives world (who can be very critical of the metadata frameworks based on it)

Key links:

- OAIS Reference Model: http://public.ccsds.org/publications/archive/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/