

An overview of the Reference Model for an Open Archival Information System (OAIS)

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

CETIS Metadata and Digital Repository SIG Meeting,
Higher Education Academy, York, 1 March 2006

<http://www.ukoln.ac.uk/>



Presentation outline

- The OAIS Reference Model
 - Background
 - Definitions, high level concepts, mandatory responsibilities
 - Functional Model
 - Information Model
- Implementing the model
- Conclusions from UKDA and TNA assessment



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



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>



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



OAIS definitions (1)

- Provides definitions of terms, e.g.:
 - **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"
 - **Designated Community** - the community of stakeholders and users that the OAIS serves
 - **Knowledge Base** - a set of information, incorporated by a user or system, that allows that user or system to understand the received information

OAIS definitions (2)

- **Information Object** - Data Object + Representation Information
- **Representation Information** - any information required to render, interpret and understand digital data
- **Information Package** - Conceptual linking of Content Information + Preservation Description Information + Packaging Information (Submission, Archival and Dissemination Information Packages)
- **Preservation Description Information** - information (metadata) about Provenance, Context, Reference, Fixity information



<http://www.ukoln.ac.uk/>

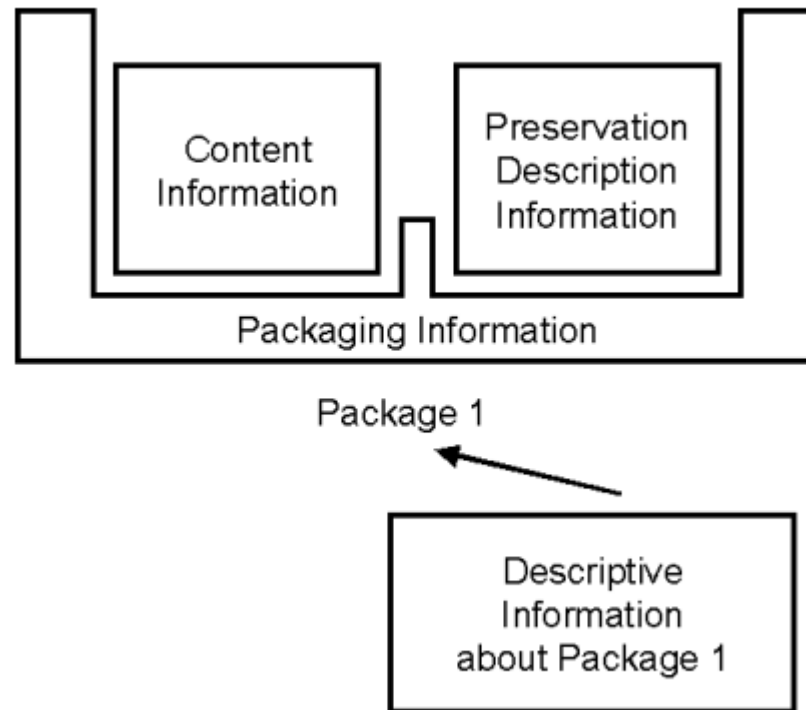
CETIS Metadata and Digital Repository SIG, York, 1 March 2006



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 can be *independently understandable*, i.e. understood without the assistance of those who first produced it
- Following documented policies and procedures
- Making the preserved information available



<http://www.ukoln.ac.uk/>

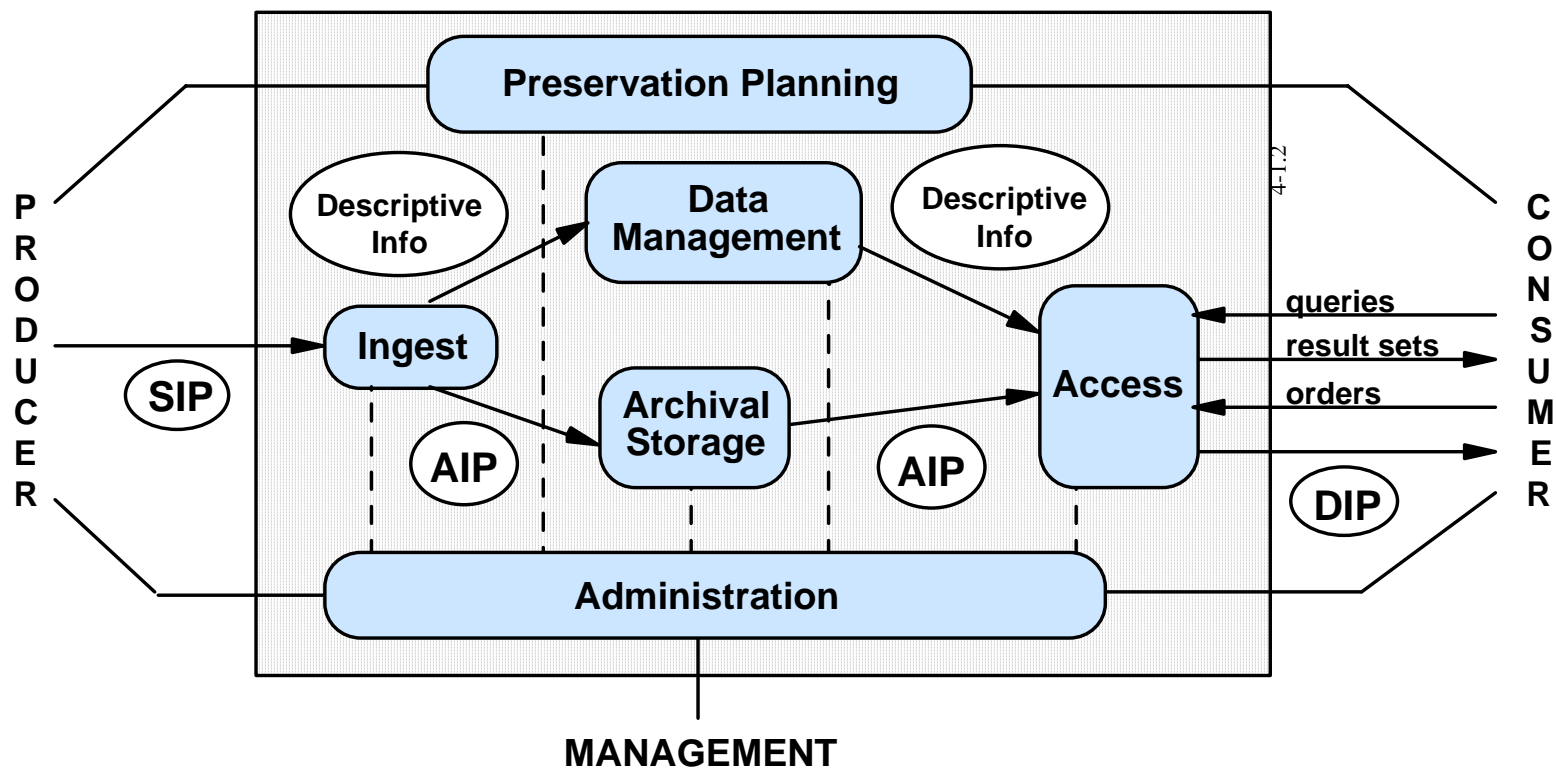
CETIS Metadata and Digital Repository SIG, York, 1 March 2006



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 Functional Entities (1)

- **Ingest** - services and functions that accept SIPs from Producers; prepares AIPs for storage, and ensures that AIPs and their supporting Descriptive Information become established within the OAIS
- **Archival Storage** - services and functions used for the storage and retrieval of AIPs
- **Data Management** -services and functions for populating, maintaining, and accessing a wide variety of information



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



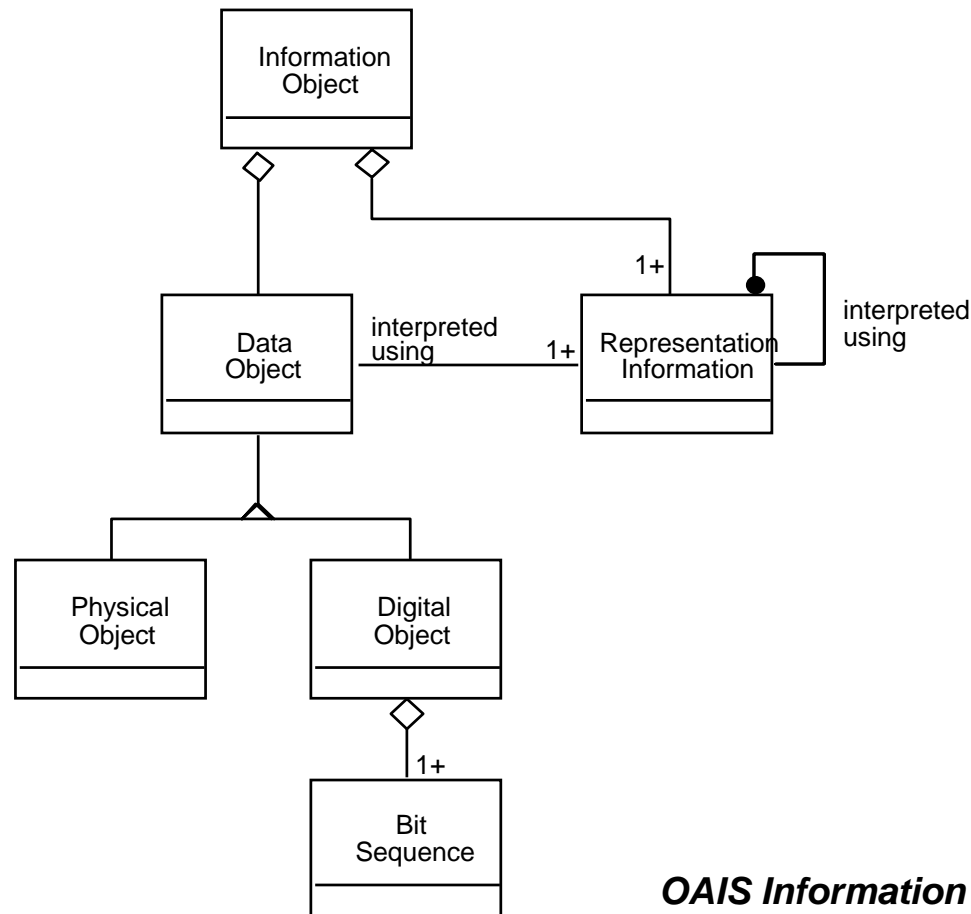
OAIS Functional Entities (2)

- **Administration** - services and functions needed to control the operation of the other OAIS functional entities on a day-to-day basis
- **Preservation Planning** - services and functions for monitoring the OAIS environment and ensuring that content remains accessible to the Designated Community
- **Access** - services and functions which make the archival information holdings and related services visible to Consumers

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)



OAIS Information Object (Figure 4-10)

OAIS Information Model (3)

- Representation Information:
 - *Any* information required to render, interpret and understand digital data (includes file formats, software, algorithms, standards, semantic information etc.)
 - Representation Information is recursive in nature
 - Essential that Representation Information itself is curated and preserved to maintain access to (render and interpret) digital data
 - e.g. Format registries (GDFR, PRONOM)

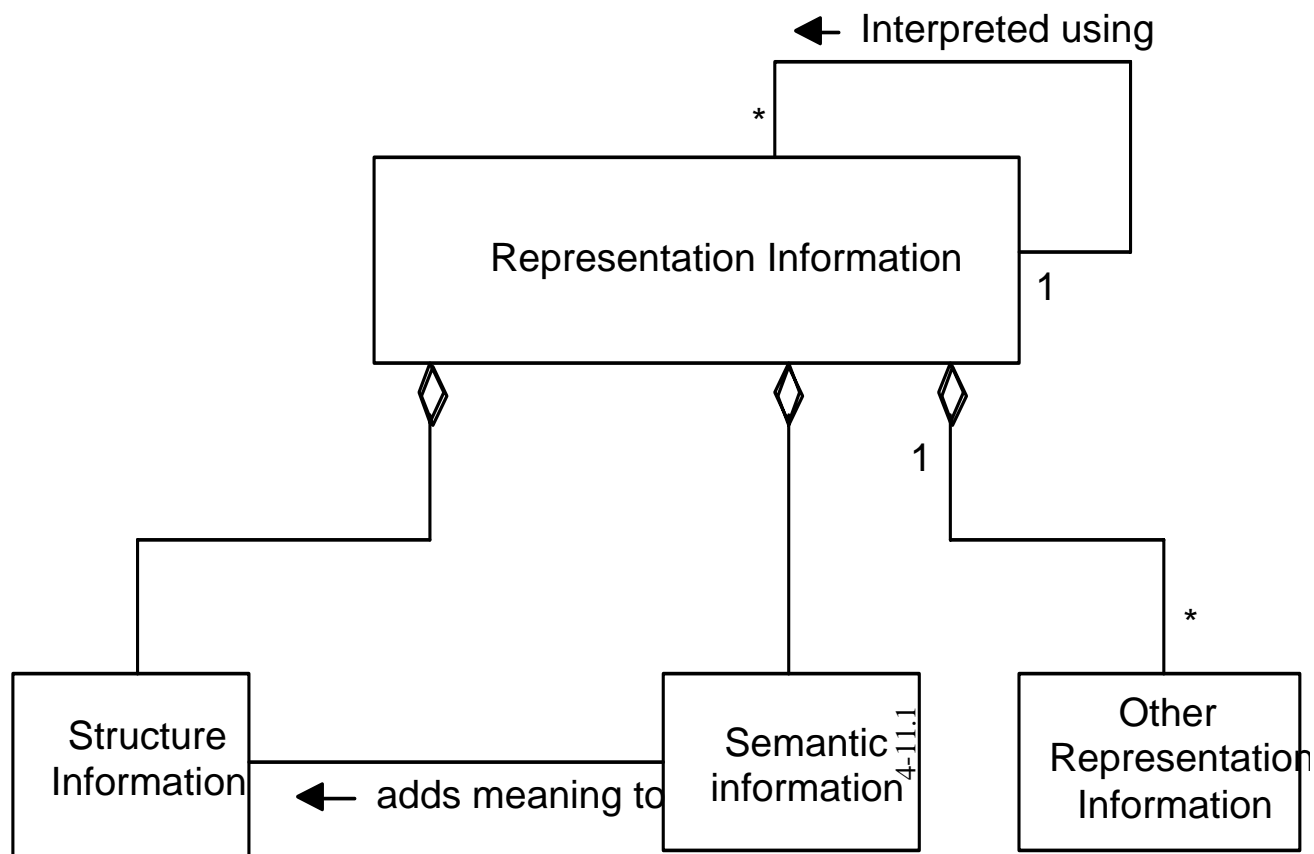


<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



OAIS Information Model (4)



OAIS Representation Information Object (Figure 4-11)

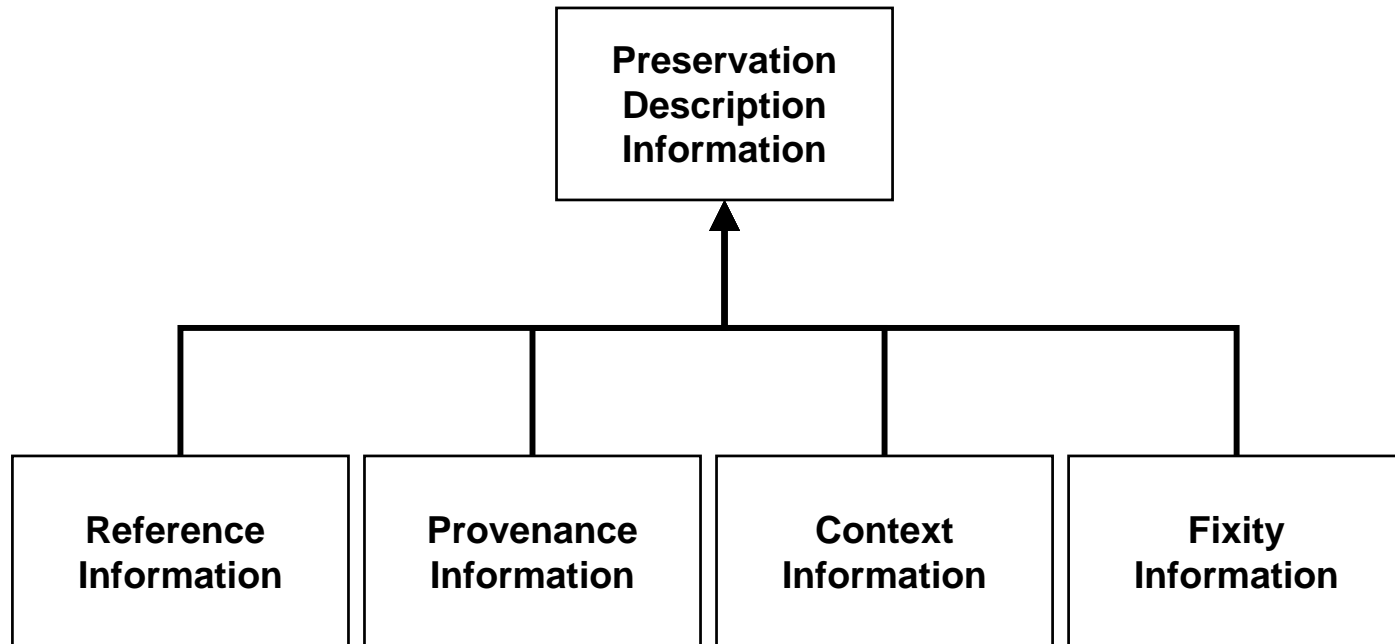
OAIS Information Model (5)

- 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 (6)

- 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
 - In part based on categories discussed in CPA/RLG report: *Preserving Digital Information* (1996)

OAIS Information Model (7)



PDI Preservation Description Information (Figure 4-16)

OAIS Information Model (8)

- **Fixity** - supporting data integrity checking mechanisms
- **Reference** - for supporting identification and location over time
- **Context** - documenting the relationship of the Content Information to its environment
- **Provenance** - documents the history of the Content Information

OAIS Information Model (9)

- Also defines:
 - Archival Information Units and Archival Information Collections
 - Recognises the complexity some some objects, addresses granularity
 - Information Package transformations
 - For Ingest and Access

OAIS - other perspectives

– Preservation

- Migration, e.g refreshment, replication, repackaging, transformation
- Preservation of look and feel (e.g., emulation, virtual machines)

– Archive interoperability

- Interaction between OAIS archives (e.g., co-operating and federated archives)

– Examples of existing archives (annex)



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



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



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Implementing OAIS (2)

- ISO 14721:2003, published in early 2003 - follows the Recommendation made available by the CCSDS
- However, earlier versions of the model made available by the CCSDS informed implementations long before its publication by ISO
- Three broad areas of influence:
 - Preservation metadata schemas
 - As a basis for further process modelling
 - Architecture and system design
 - Conformance criteria for repositories

Implementing OAIS - metadata (1)

- 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) ...



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Implementing OAIS - metadata (2)

– PREMIS Data Dictionary

- OAIS remains the conceptual foundation (but there are now some differences in terminology)
- Preservation metadata = "the information a repository uses to support the digital preservation process"
- The PREMIS Data Dictionary defines metadata that supports "maintaining viability, renderability, understandability, authenticity, and identity in a preservation context."
- Core metadata = "things that most working repositories are likely to need to know in order to support digital preservation."



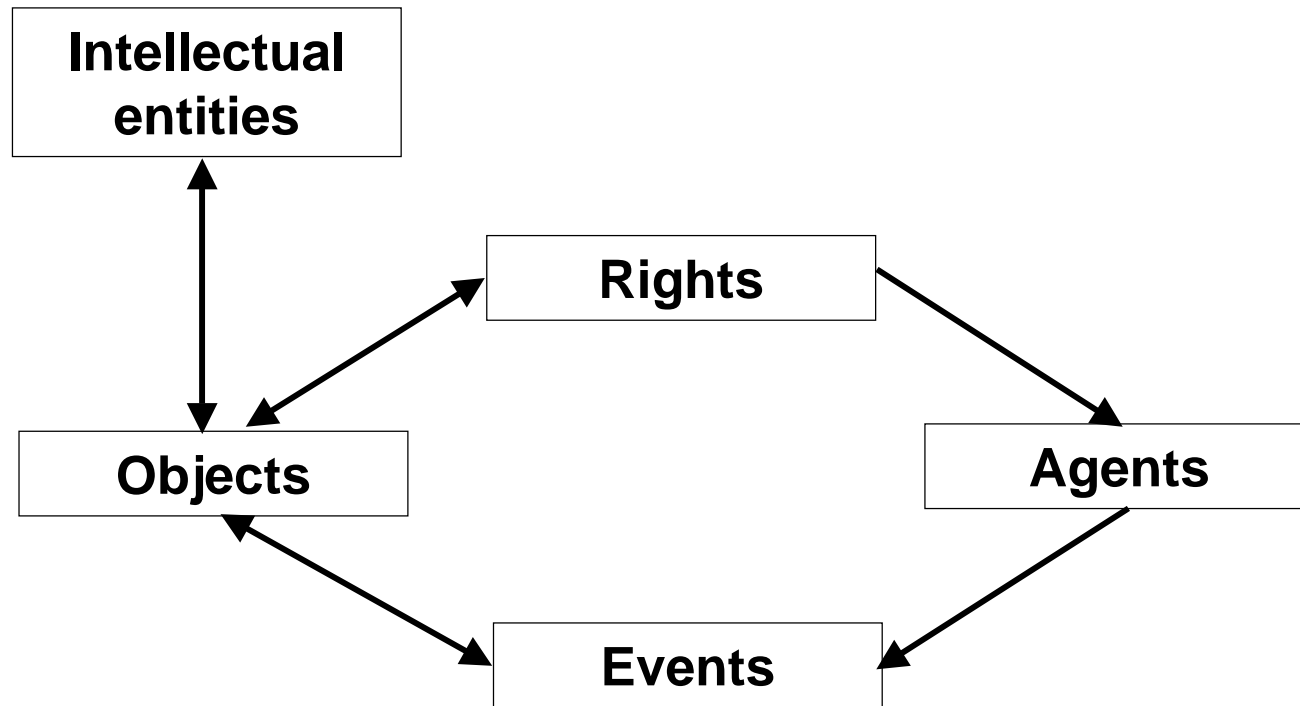
<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Implementing OAIS - metadata (3)

PREMIS Data Model



Implementing OAIS - modelling

- InterPARES Preservation Task Force
 - Preserve Electronic Records model
 - Modelled the specific processes and functions involved with preserving electronic records
 - "... a specification of an OAIS for the specific classes of information objects comprising electronic records and archival aggregates of such records"
 - <http://www.interpares.org/>



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



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, The National Archives
 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)



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Implementing OAIS - systems (2)

– Examples:

- 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 networks
- DCC Representation Information Registry/Repository, DSpace, KB e-Depot, ...



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Implementing OAIS - conformance (1)

- Many repositories or preservation tools claim OAIS influence or 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 (reprise):
 - 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.



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



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

UKDA and TNA study (1)

- JISC project:

- Mapping to functional and information models
- Beedham, H., *et al.*, (2005). *Assessment of UKDA and TNA Compliance with OAIS and METS Standards*. Available
<http://www.data-archive.ac.uk/news/publications/oaismets.pdf>

- Some conclusions:

- Noted that there was no existing methodology for testing OAIS compliance
 - » Recommended the production of guidelines or manual



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



UKDA and TNA study (2)

– Conclusions (continued)

- The OAIS Mandatory Responsibilities are carried out by almost any archive
- The OAIS Designated Community concept assumes a identifiable and relatively homogenous user community; this is not the case for either UKDA or TNA
- The relationship between AIPs and DIPs needs clarification
- The OAIS Administration function may be difficult for small archives to fulfil adequately

UKDA and TNA study (3)

- Conclusions (continued):
 - Model not scalable - report proposes an 'OAIS Lite'
 - Information categories are too general to allow mapping of metadata elements from other schemas (p. 70)
 - But ... OAIS terminology was useful to support communication between UKDA and TNA

Some personal comments (1)

- Conformance with the OAIS model is often claimed by digital preservation efforts (e.g. DSpace, METS, LOCKSS) - 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., UK Data Archive and TNA study, BADC
- 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 now be used as a starting point for conformance



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Some personal comments (2)

- Sometimes interpreters of the model seem reluctant to acknowledge that it could be improved
- There may be a need for periodic revision, some clarification of definitions; possibly more feedback from the archives world



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006



Key links:

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



<http://www.ukoln.ac.uk/>

CETIS Metadata and Digital Repository SIG, York, 1 March 2006

