

---

# ***Reference Model for an Open Archival Information Systems (OAIS): Overview and Current Status***

***Alan Wood/A.E.Wood & Erickson/Lockheed Martin***

***Don Sawyer/NASA/GSFC***

***Lou Reich/CSC***

***16-June-2001***

***American Library Association Annual Conference***

# ***What is a Reference Model?***

---

## ■ **A framework**

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

## ■ **A reference model**

- **is based on a small number of unifying concepts**
- **is an abstraction of the key concepts, their relationships, and their interfaces both to each other and to the external environment**
- **may be used as a basis for education and explaining standards to a non-specialist.**

# ***What was the Motivation?***

---

- **Agencies and organizations have a significant stewardship responsibility for the digital information obtained from their programs**
  - 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 and must be readily transportable from archive to archive**

# ***Why a 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
- **Developing ISO-level archive implementation standards would be wasted effort without an agreed framework among users and archives**
- **First need an archive reference model**

# ***Organizational Approach***

- **Organize US contribution under a framework with NASA lead**
  - Establish liaison with Federal Geographic Data Committee (FGDC) and National Archives and Records Administration (NARA)
  - Agency archives and users must be represented in this process
- **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
- **Broad international workshops also held**
  - Britain, Germany, Italy, France, Spain
- **Issue resolution at ISO/Consultative Committee for Space Data Systems 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 Geomatics
- **Define what is meant by ‘archiving of data’**
- **Break ‘archiving’ into a few functional areas (e.g., ingest, storage, access, and preservation planning)**
- **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
  - Unified Modeling Language (UML) 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
- **Already widely adopted as starting point in digital preservation efforts**
  - Digital libraries (e.g., Netherlands National Library)
  - Traditional archives (e.g., US National Archives)
  - Scientific data centers (e.g., National Space Science Data Center, NASA Space Life Sciences Data Archives)
  - Commercial Organizations (e.g., Aerospace Industries Association preservation working team)

# Reference Model Status

- Completed CCSDS Red Book review in November 2000
  - [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html)
- Completed ISO Draft International Standard (DIS) review
  - Same content as CCSDS Red Book
- Comments received from several organizations
  - Update underway
    - [http://ssdoo.gsfc.nasa.gov/nost/isoas/us20/650x0\\_010510.pdf](http://ssdoo.gsfc.nasa.gov/nost/isoas/us20/650x0_010510.pdf)
  - Major impact is to highlight the preservation planning function in the functional model
- Plan to have a new ISO Draft International Standard 2 month review starting July 2001.
- Assuming only minimal delays this should produce a final standard in the Fall, 2001

# ***Reference Model for an Open Archival Information System***

## ***Brief Technical Overview***

# ***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

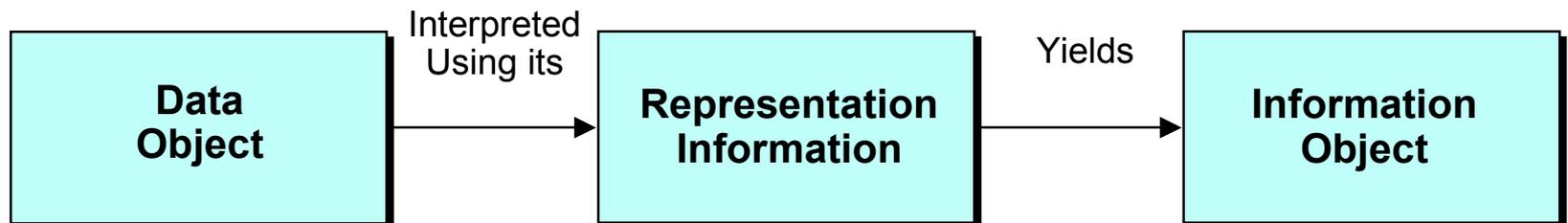
# ***Purpose, Scope, and Applicability***

---

- **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**

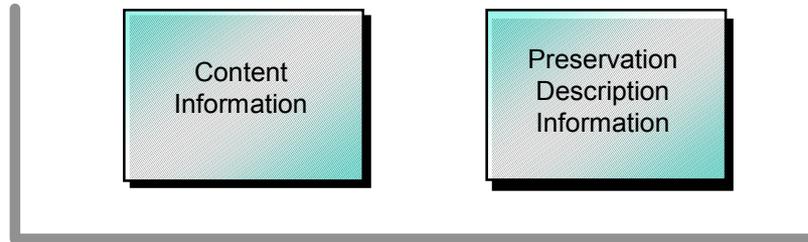
# OAIS Information Definition

- 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**



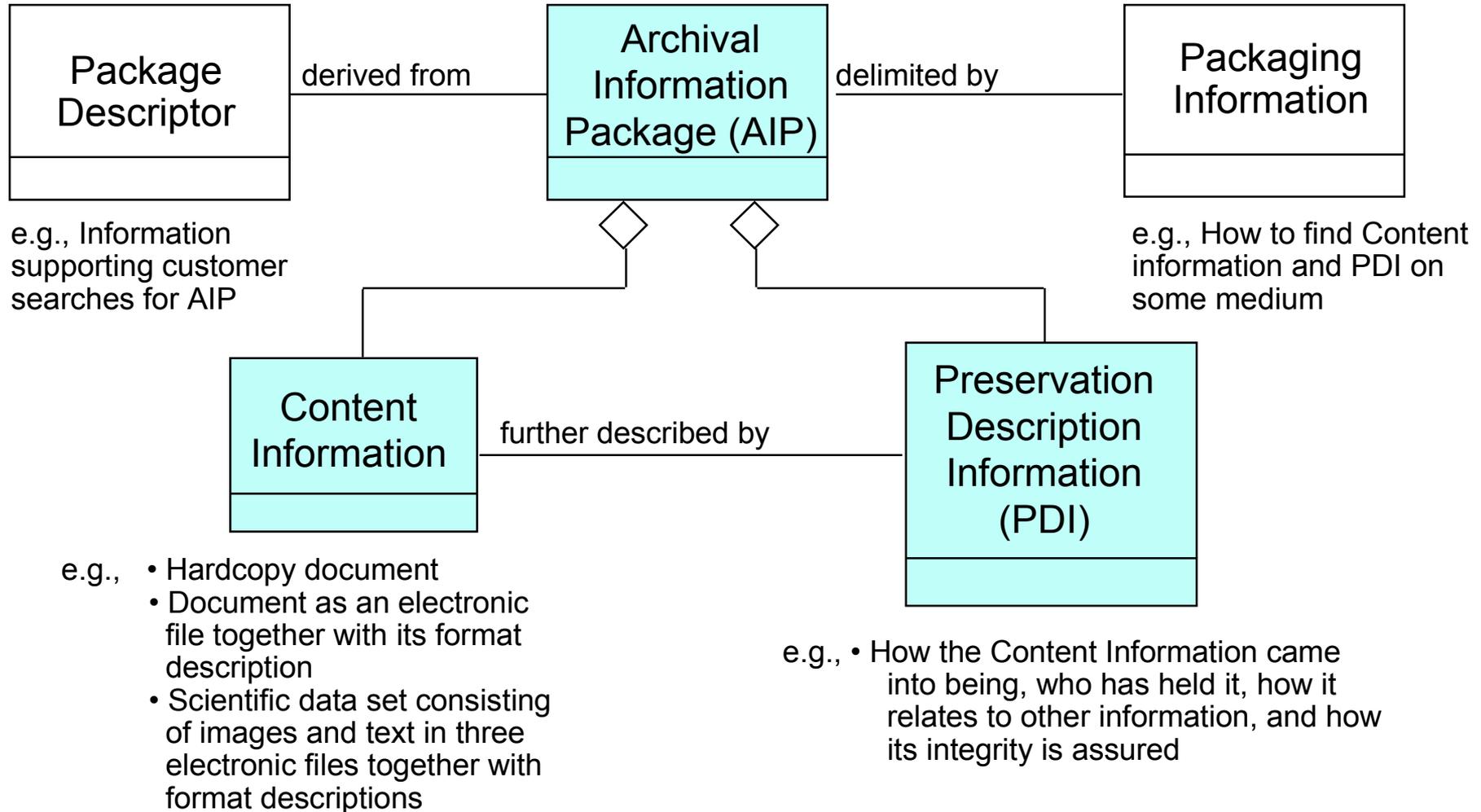
# *Information Package Definition*

---

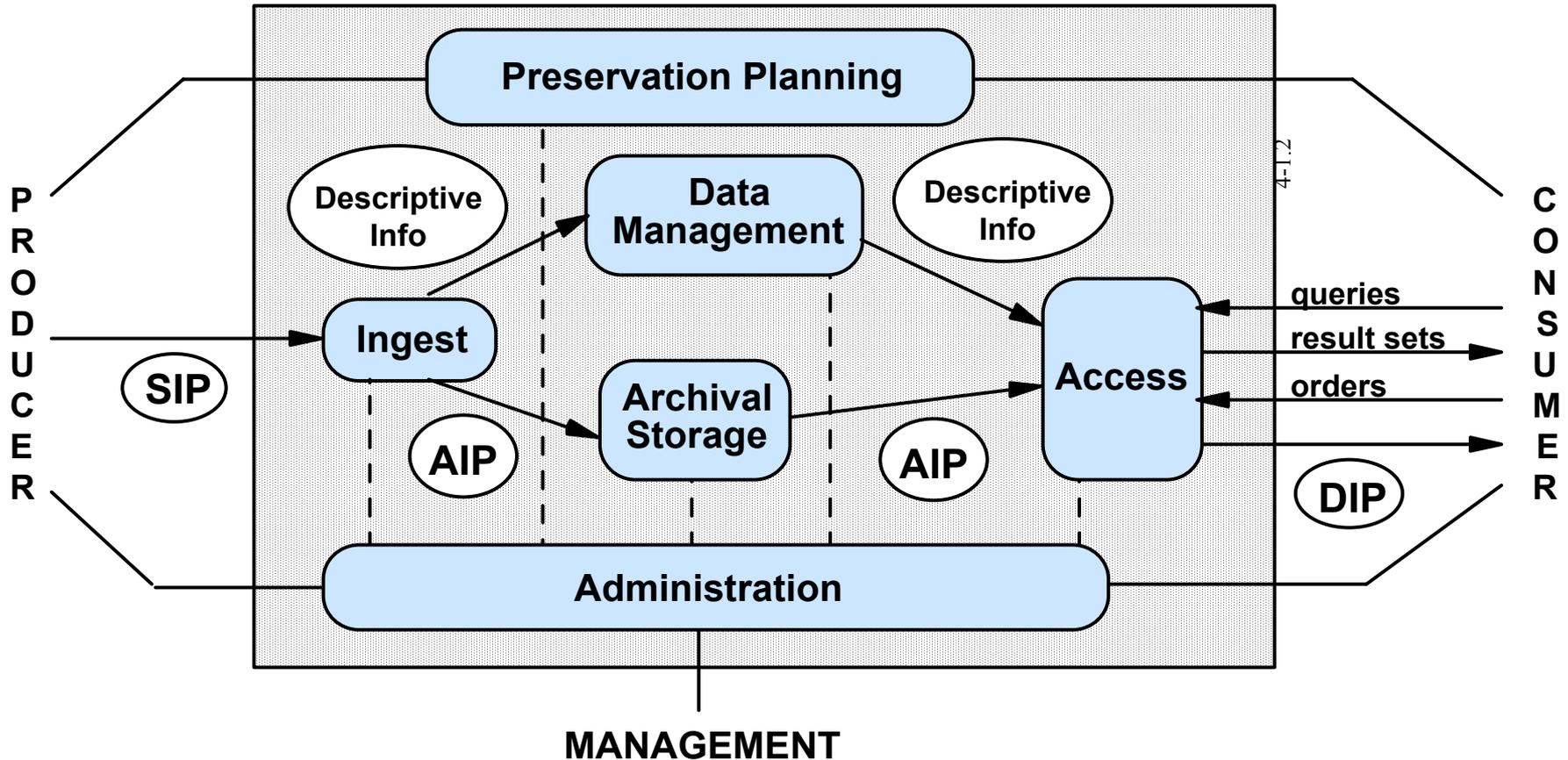


- An **Information Package** is a conceptual container of two types of information called **Content Information** and **Preservation Description Information (PDI)**

# OAIS Archival Information Package



# OAIS Functional Entities



SIP = Submission Information Package  
AIP = Archival Information Package  
DIP = Dissemination Information Package

# *Functional Entities In An OAIS*

- **Ingest:** This entity provides the services and functions to accept Submission Information Packages (SIPs) from Producers and prepare the contents for storage and management within the archive
- **Archival Storage:** This entity provides the services and functions for the storage, maintenance and retrieval of Archival Information Packages
- **Data Management:** This entity provides the services and functions for populating, maintaining, and accessing both descriptive information which identifies and documents archive holdings and internal archive administrative data.
- **Administration:** This entity manages the overall operation of the archive system
- **Preservation Planning:** This entity monitors the environment of the OAIS and provides recommendations to ensure that the information stored in the OAIS remain accessible to the Designated User Community over the long term even if the original computing environment becomes obsolete.
- **Access:** This entity supports consumers in determining the existence, description, location and availability of information stored in the OAIS and allowing consumers to request and receive information products

# *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**

---

# ***Some Applications***

# ***Basis of Systems Architecture in Digital Libraries and Scientific Archives***

- NEDLIB (Networked European Deposit Library) effort used OAIS Reference Model as a basis for the design and architecture of **Deposit System for Electronic Publications (DSEP)**
- **National Library of Australia** using it as basis for their implementation
- **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
- **NSSDC (National Space Science Data Center )** is evolving their archive using OAIS RM as a basis for a new architecture
- **SIPAD**: French space agency plasma physics archive used the OAIS as a basis for design

# ***Enhanced Communications and Productivity among varied Communities***

---

- **National Archives and Records Administration 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 several weeks of effort in the specification of the task**
- **Similar experiences between:**
  - **National Library of France and French space agency (CNES) representatives**
  - **National Center for Supercomputer Applications HDF format developers and DNA researchers**
  - **NASA Life Sciences Archive and micro-gravity researchers**
  - **United States Department of Agriculture and digital preservation experts**

# ***More OAIS Accomplishments***

---

- **Royal Library of the Netherlands (RLN)**
  - OAIS mandated in their implementation RFP
  - IBM implementing OAIS-based system for RLN (£5M project)
- **France setting 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
- **OAIS likely to be used by International Council for Scientific Unions (ICSU) as basis for study on long-term preservation**

# Reference URLs

---

- **May 10, 2001 OAIS draft**
  - [http://ssdoo.gsfc.nasa.gov/nost/isoas/us20/650x0\\_010510.pdf](http://ssdoo.gsfc.nasa.gov/nost/isoas/us20/650x0_010510.pdf)
- **ISO Archive Standards Overview Web site**
  - <http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>
- **Lavoie, Brian. "Meeting the challenges of digital preservation: the OAIS reference model". OCLC Newsletter. No. 243. January/February 2000. Pages 26-30.**  
***\*An excellent overview of the OAIS RM and Workshops.\****
  - <http://www2.oclc.org/oclc/pdf/news243.pdf>