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

Lou Reich/NASA/CSC

Don Sawyer/NASA/GSFC

19-Oct-2001

Digital Curation Workshop

London

Organizational Backround

- **National Space Science Data Center**
  - NASA's first digital archive
  - Experienced many technology changes since 1966
- **Consultative Committee for Space Data Systems**
  - International group of space agencies
  - Developed variety of science discipline- independent standards
  - Became working body for an ISO TC 20/ SC 13 about 1990
- **ISO suggested that SC 13 should develop archive standards**
  - Address data used in conjunction with space missions
  - Address intermediate and indefinite long term storage of digital data

Response

- **Response to Consultative Committee for Space Data Systems (CCSDS) and ISO TC 20/SC 13**
  - **No framework widely recognized for developing specific digital archive standards**

  - **Begin by developing a 'Reference Model' to establish common terms and concepts**

  - **Ensure broad participation, including traditional archives**

    (Not restricted to space communities; all participation is welcome!)

  - **Focus on data in electronic forms, but recognize that other forms exist in most archives**

  - **Follow up with additional archive standards efforts as appropriate**

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**
- **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**
  - **Information must be readily transportable from archive to archive**

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**
  - **UK and France**
  - **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 Geomantics**
- **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)**
  - — **Commercial Organizations (e.g., Aerospace Industries Association preservation working team)**

Reference Model Status

- **Completed CCSDS Red Book review in November 2000**
- **Completed ISO Draft International Standard (DIS) review**
  - — **Same content as CCSDS Red Book**
- **Comments were received from several organizations**
  - — **Issues discussed and resolved at the November 2000 and May 2001 ISO Archiving Workshop**
  - — **Major impact is to highlight the preservation planning function in the functional model**
- **A new version of the OAIS Reference Model was delivered to the ISO and CCSDS Secretariats in July 2001 for a two month public review period http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf**
- **Assuming only minimal delays this should produce a final standard in the Fall, 2001**

Reference Model for an Open Archival Information System

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 Responsibilities

- **Negotiates and accepts Information Packages from information producers**

- **Obtains sufficient control to ensure long-term preservation**

- **Determines which communities (designated) need to be able to understand the preserved information**

- **Ensures the information to be preserved is independently understandable to the Designated Communities**

- **Follows documented policies and procedures which ensure the information is preserved against all reasonable contingencies**

- **Makes the preserved information available to the Designated Communities in forms understandable to those communities**

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**

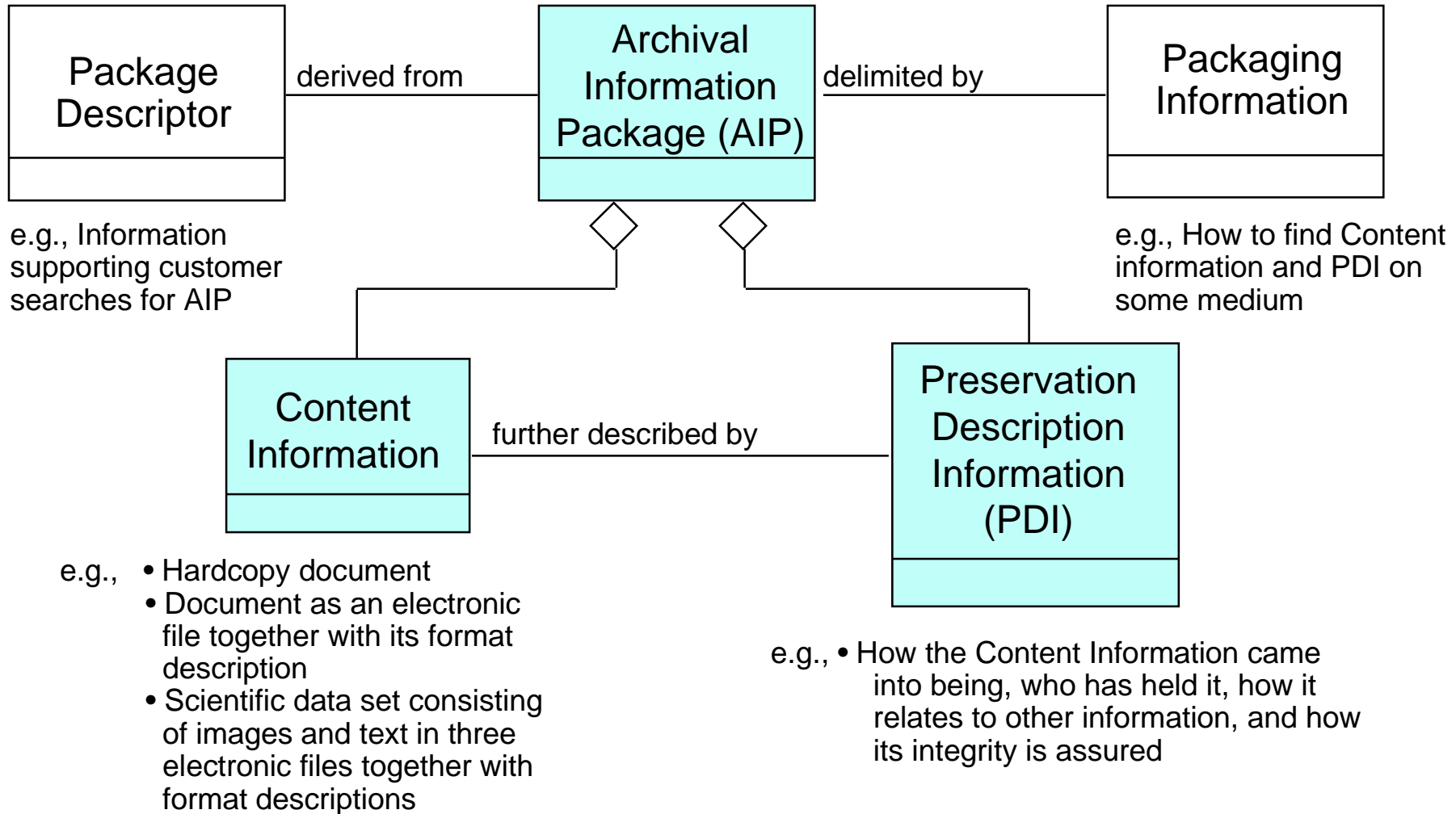| Data Object | Interpreted Using its → | Representation Information | Yields → | Information Object |
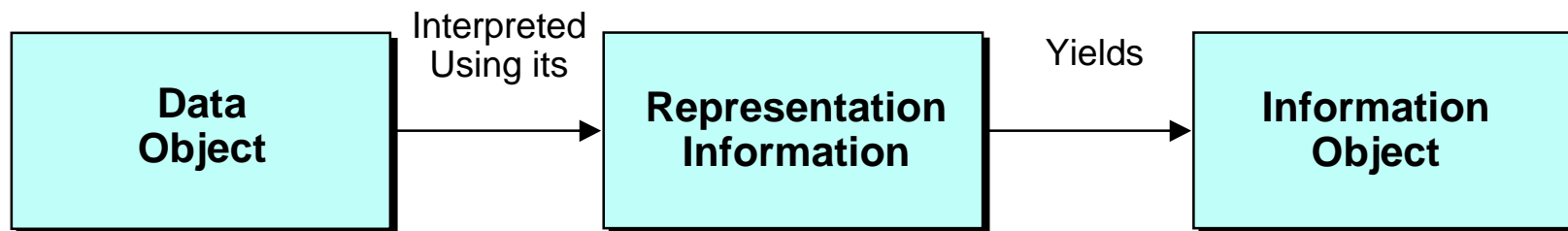
Information Package Definition
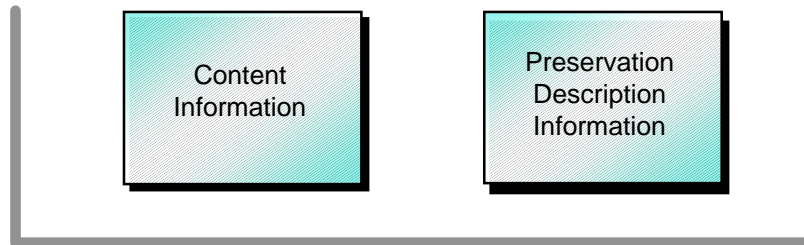
- **An Information Package is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI)**

OAIS Archival Information Package

Package Descriptor	derived from	Archival Information Package (AIP)	delimited by	Packaging Information

e.g., Information supporting customer searches for AIP

e.g., How to find Content information and PDI on some medium

Content Information — further described by — **Preservation Description Information (PDI)**

e.g., • Hardcopy document
  • Document as an electronic file together with its format description
  • Scientific data set consisting of images and text in three electronic files together with format descriptions

e.g., • How the Content Information came into being, who has held it, how it relates to other information, and how its integrity is assured
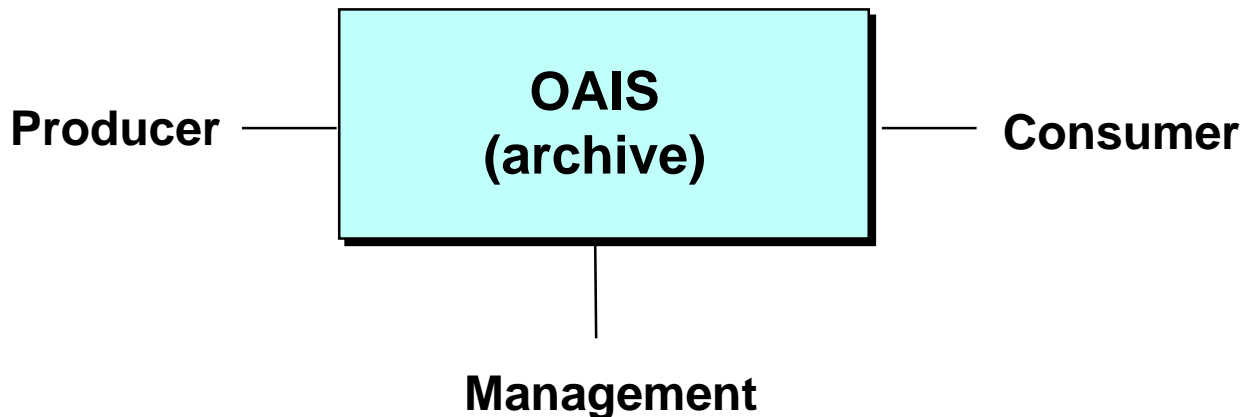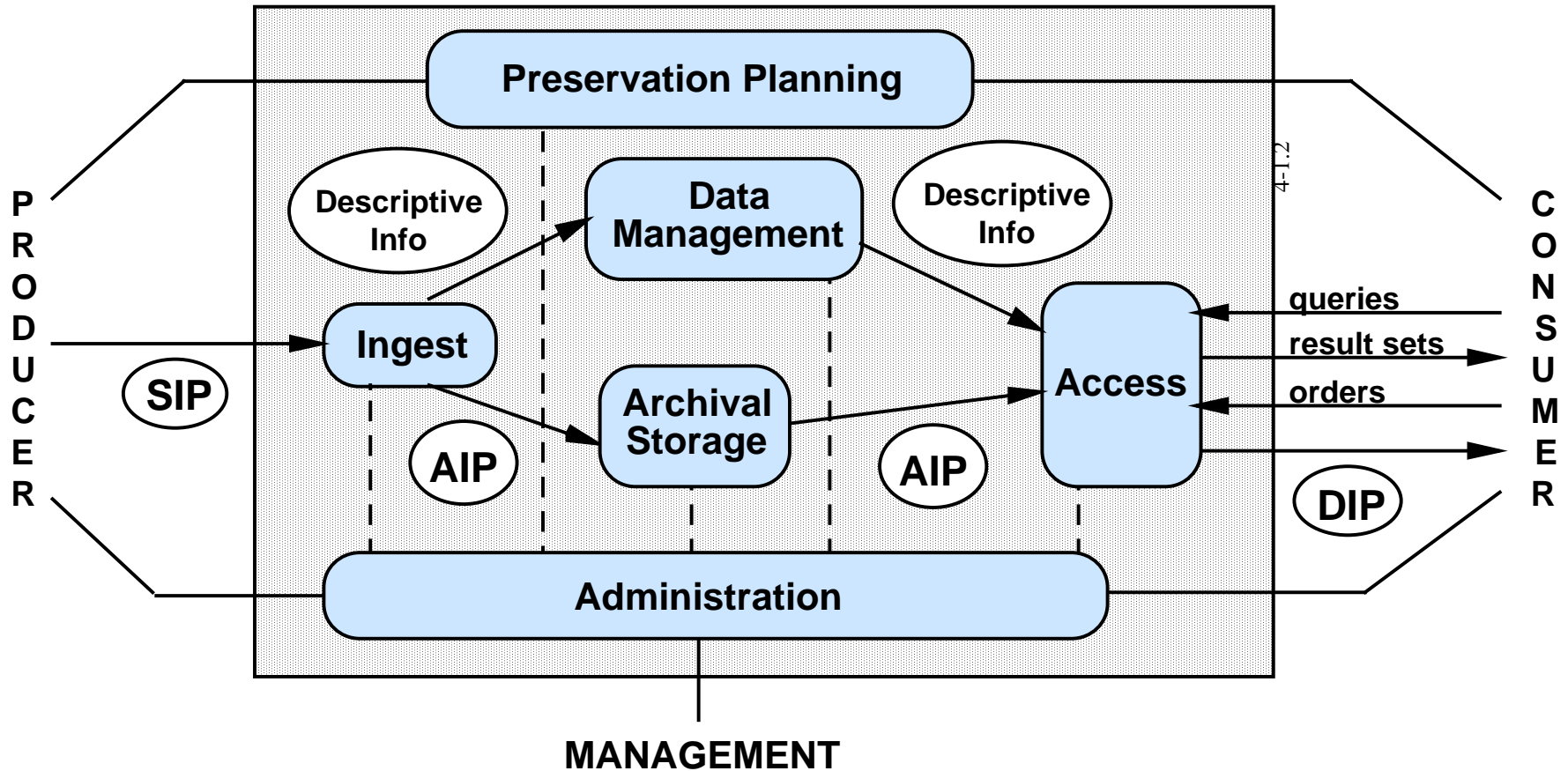
Information Package Variants

- **Submission Information Package**
  - **Negotiated between Producer and OAIS**
  - **Sent to OAIS by a Producer**
- **Archival Information Package**
  - **Information Package used for preservation**
  - **Includes complete set of Preservation Description Information for the Content Information**
- **Dissemination Information Package**
  - **Includes part or all of one or more Archival Information Packages**
  - **Sent to a Consumer by the OAIS**

Model View of an OAIS Environment

- **Producer is the role played by those persons, or 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 those persons, or client systems, who interact with OAIS services to find and acquire preserved information of interest**

Producer ——— | **OAIS (archive)** | ——— **Consumer**

Management

OAIS Functional Entities

MANAGEMENT

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

Preservation Planning
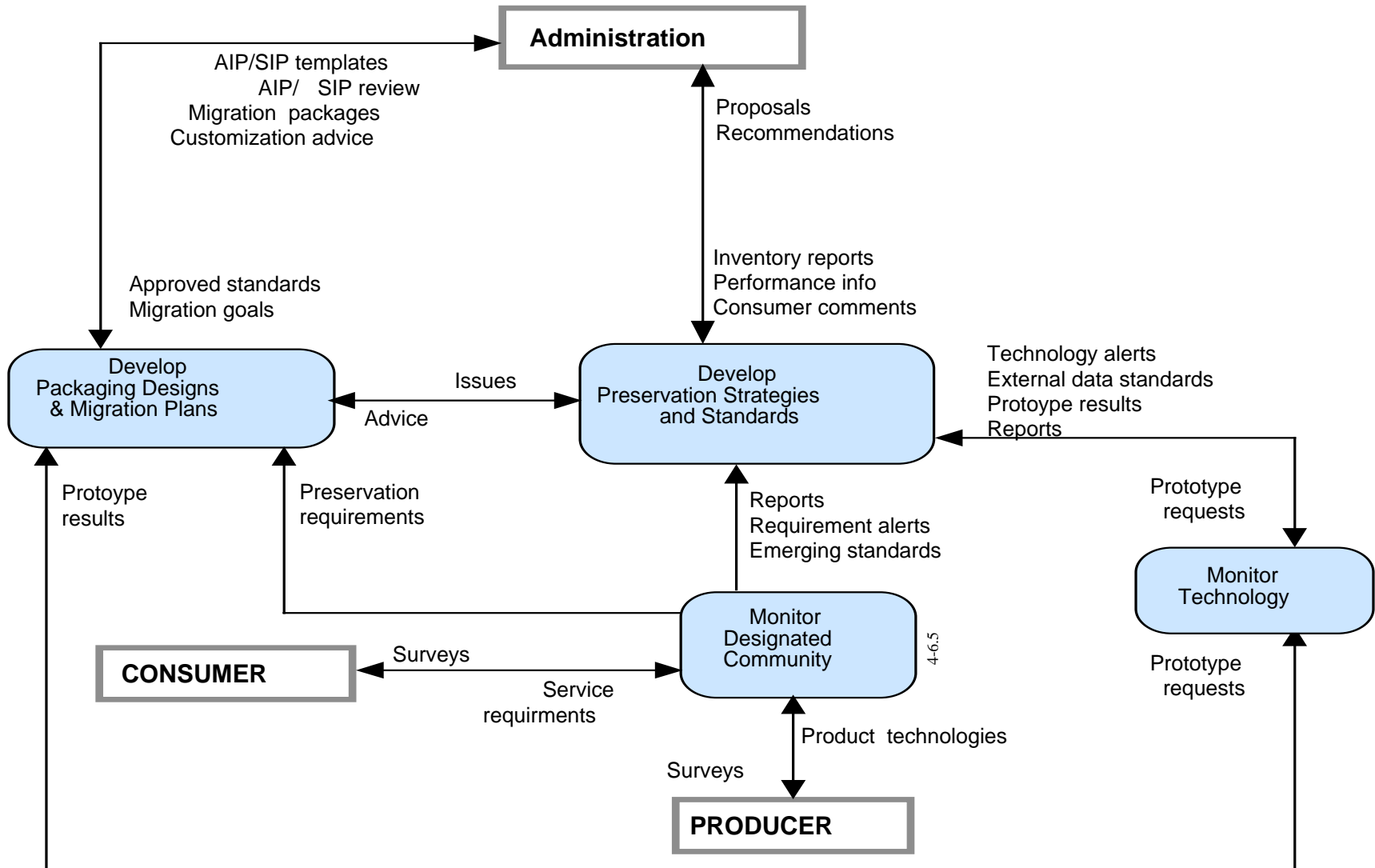
Reference Model Summary

- **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**
  - **Life Sciences Archive developer 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**

Comments from the RLG/OCLC Preservation Commons Papers

Attributes of a Trusted Repository for Digital Materials,

Recommendation 13: Stakeholders in the research repository community should formally endorse the Open Archival Information Systems Reference Model, actively supporting and advocating it as a standard for digital repository services. Specific guidance and support for implementation will be needed and could take the form of advisory services offered by expert organizations.

Recommendation 14: The OAIS model will be more useful to libraries and archives as many of its functions can be automated. Work will be needed to consider existing systems and provide recommendations and specifications for further development work or enhancements for automation.

Summary: A review of the Open Archival Information Systems model provides insight into the functions necessary for the establishment of a long-term digital repository. Its adoption by libraries and repositories is increasing, due mainly to continued involvement of the library and archives communities in the development of the model itself. Although work is still to be done on many of the standards necessary for effective distributed implementation of the model, OAIS provides a critical framework for establishing or enhancing digital archiving services.

OAIS Appraisal from the RLG/OCLC Preservation Commons Papers

Preservation Metadata for Digital Objects: A Review of the State of the Art

OAIS Reference Model: **The preservation metadata approaches reviewed in this paper share the characteristic of being informed, either explicitly or implicitly, by the OAIS framework. The CEDARS and NEDLIB approaches explicitly adopt OAIS concepts and terminology as an underlying framework for their metadata. The NLA follows the OAIS implicitly in the sense that its metadata elements can be mapped tothe OAIS information model in a straightforward way. This suggests that the OAIS may be a useful starting point for the consensus-building process undertaken by the Working Group. General agreement seems to exist that the OAIS provides a reasonable description of both the functional components of a digital archiving system, and the information requirements needed to support these components.Therefore, the prospects for adapting and/or extending the OAIS framework as a foundation for consensus on preservation metadata appear promising.**

Reference URLs

- **July 15 OAIS RM version 2**
  http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.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**