

# Digital Preservation

## OAIS Reference Model

Andreas Rauber  
Institute of Software Technology and Interactive Systems  
Vienna University of Technology  
<http://www.ifs.tuwien.ac.at/dp>

# Outline

- 
- Principles of the OAIS Model
  - Technical Overview
  - Functional Overview
  - Information Modell
  - Summary
-

# OAIS and the role of NASA

- National Space Science Data Center
  - NASA's first digital archive
  - has gone through many technology changes since 1966
- Consultative Committee for Space Data Systems
  - International group of Space Agencies
  - developed a set of standards across disciplines
  - evolved into working group ISO TC 20/ SC 13 around 1990

TC20: Aircraft and Space Vehicles

SC13: Space Data and Information Transfer Systems

# What's a reference model

- A Framework
  - to understand the relationship between significant entities in an environment
  - for the development of consistent standards or spezifikations to support this environment.
- A reference model
  - is based on a small number of unified concepts
  - is an abstraction of the core concepts, their relationships and interfaces within as well as external to the framework
  - can be used as a basis for training and to explain standards.

# OAIS

- OAIS is a reference model
- No design specification, no data model, no set of functional requirements!
- Describes elements and concepts that are relevant for a project
- Goal: determine, which parts of the reference model map to which subsystems, functions and responsibilities in a desired solution.

# OAIS Sources of Information

- Reference Model for an Open Archival Information System (OAIS), ISO 14721:2012
- Blue Book, CCSDS 650.0-B-1, January 2002
- Pink Book, CCSDS 650.0-P-1.1, August 2009
- <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf>
- Slides based on Blue Book, Pink Book and:
  - Don Sawyer, Lou Reich: ISO Reference Model for an Open Archival Information System (OAIS) Tutorial Presentation, LOC, June 13 2003
- <http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>

# Outline

- 
- Principles of the OAIS Model
  - Technical Overview
  - Functional Overview
  - Information Modell
  - Summary
-

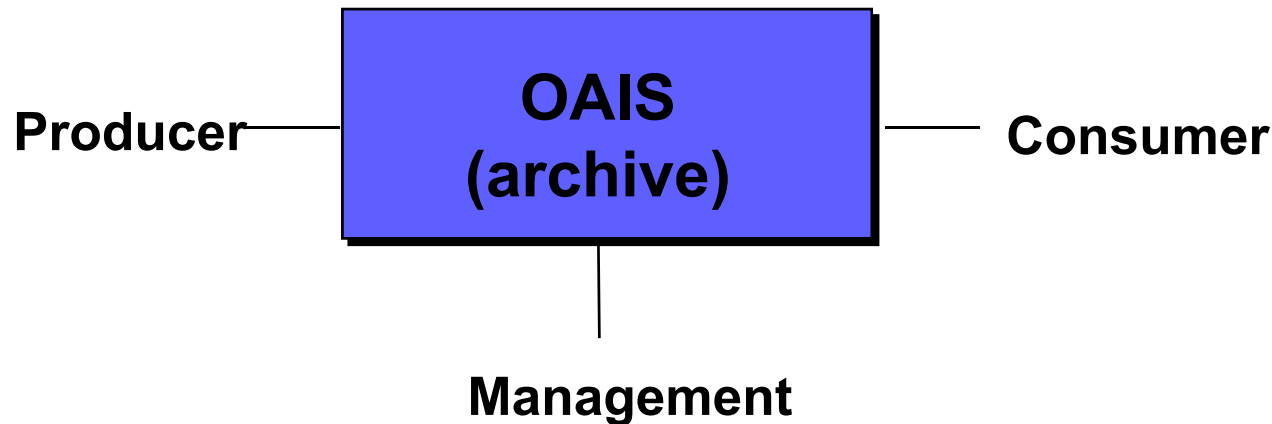
- Open
  - Reference Model standard(s) are developed using a public process and are freely available
- Information
  - Any type of knowledge that can be exchanged
  - Independent of the forms (i.e., physical or digital) used to represent the information
  - Data are the representation forms of information
- Archival Information System
  - Hardware, software, and people who are responsible for the acquisition, preservation and dissemination of the information



# 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 an implementation

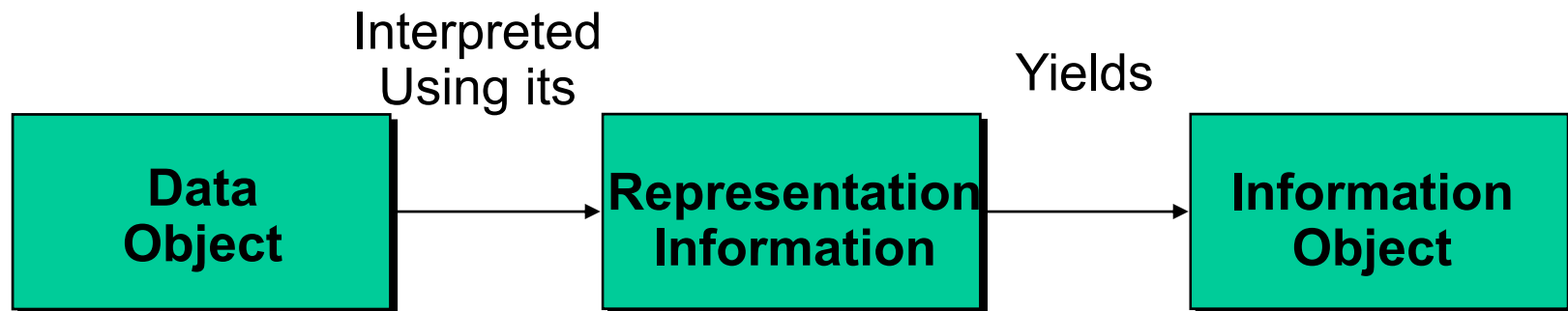
# Model View of an OAIS Environment



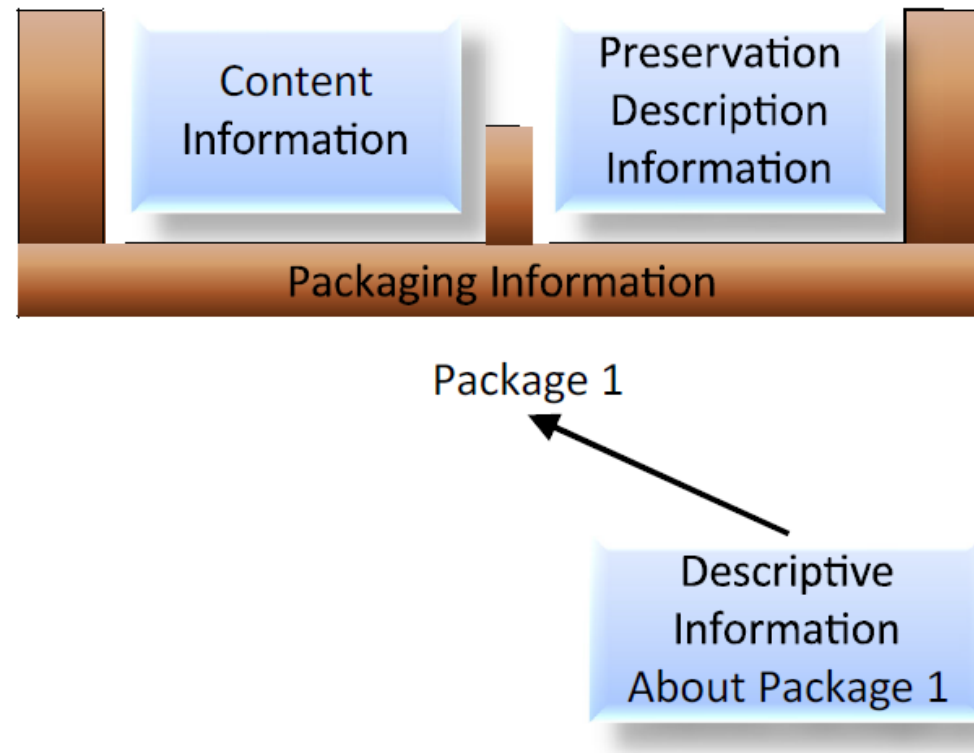
- **Producer** is the role played by those persons, or client systems, who provide the information to be preserved
- **Management** is the role played by those who set overall OAIS policy as one component in a broader policy domain
- **Consumer** is the role played by those persons, or client systems, who interact with OAIS services to find and acquire preserved information of interest

# OAIS Information Definition

- Information is always expressed (i.e., represented) by some type of data
- Data interpreted using its Representation Information yields Information
- Information Object preservation requires clear identification and understanding of the Data Object and its associated Representation Information



# Information Package Definition



- An Information Package is a conceptual container holding two types of information
  - Content Information
  - Preservation Description Information (PDI)
- Plus descriptive information

# Information Package Variants

## ■ **SIP:** Submission Information Package

- Negotiated between Producer and OAIS
- Sent to OAIS by a Producer

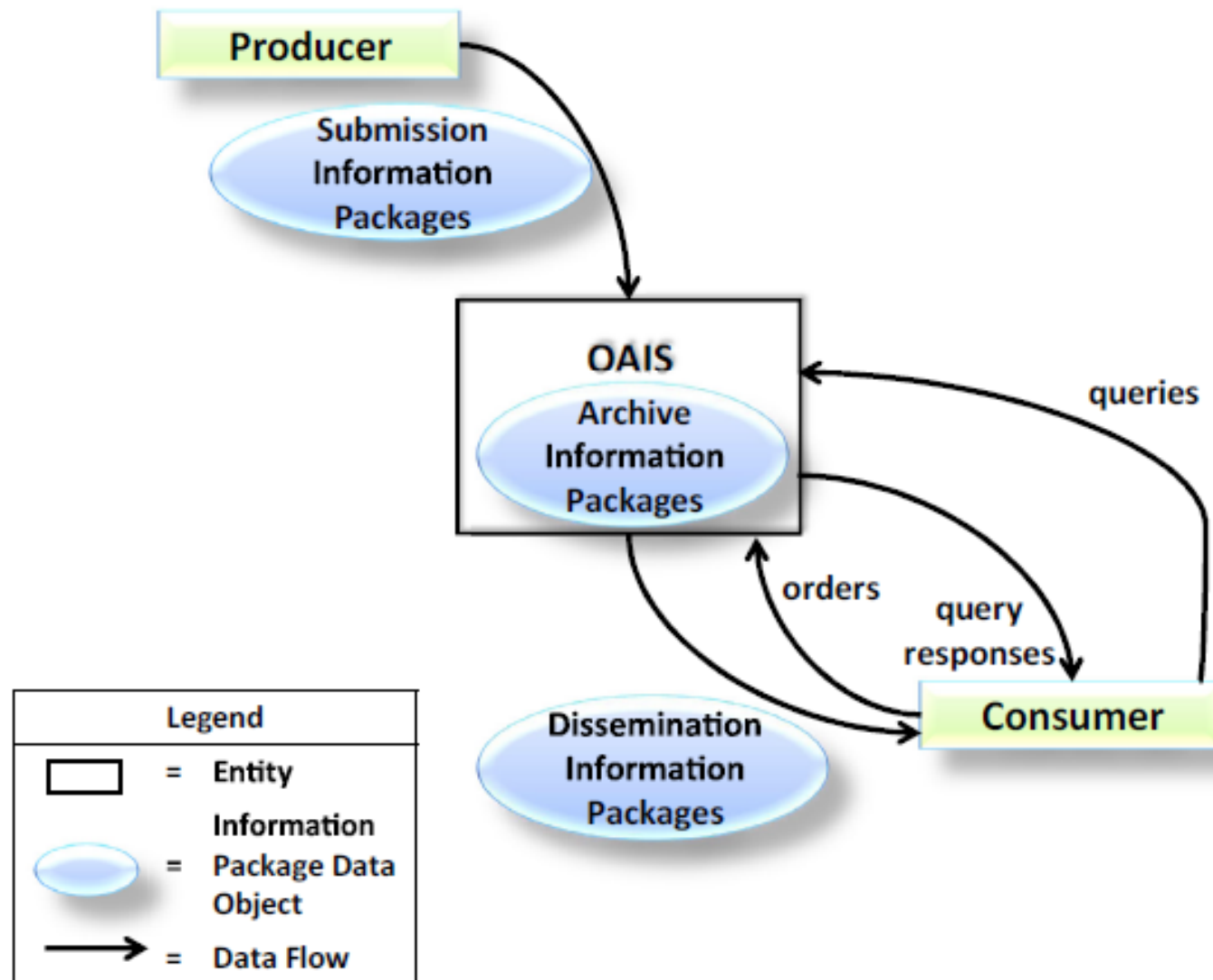
## ■ **AIP:** Archival Information Package

- Information Package used for preservation
- Includes complete set of Preservation Description Information (PDI) for the Content Information

## ■ **DIP:** Dissemination Information Package

- Includes part or all of one or more Archival Information Packages
- Sent to a Consumer by the OAIS

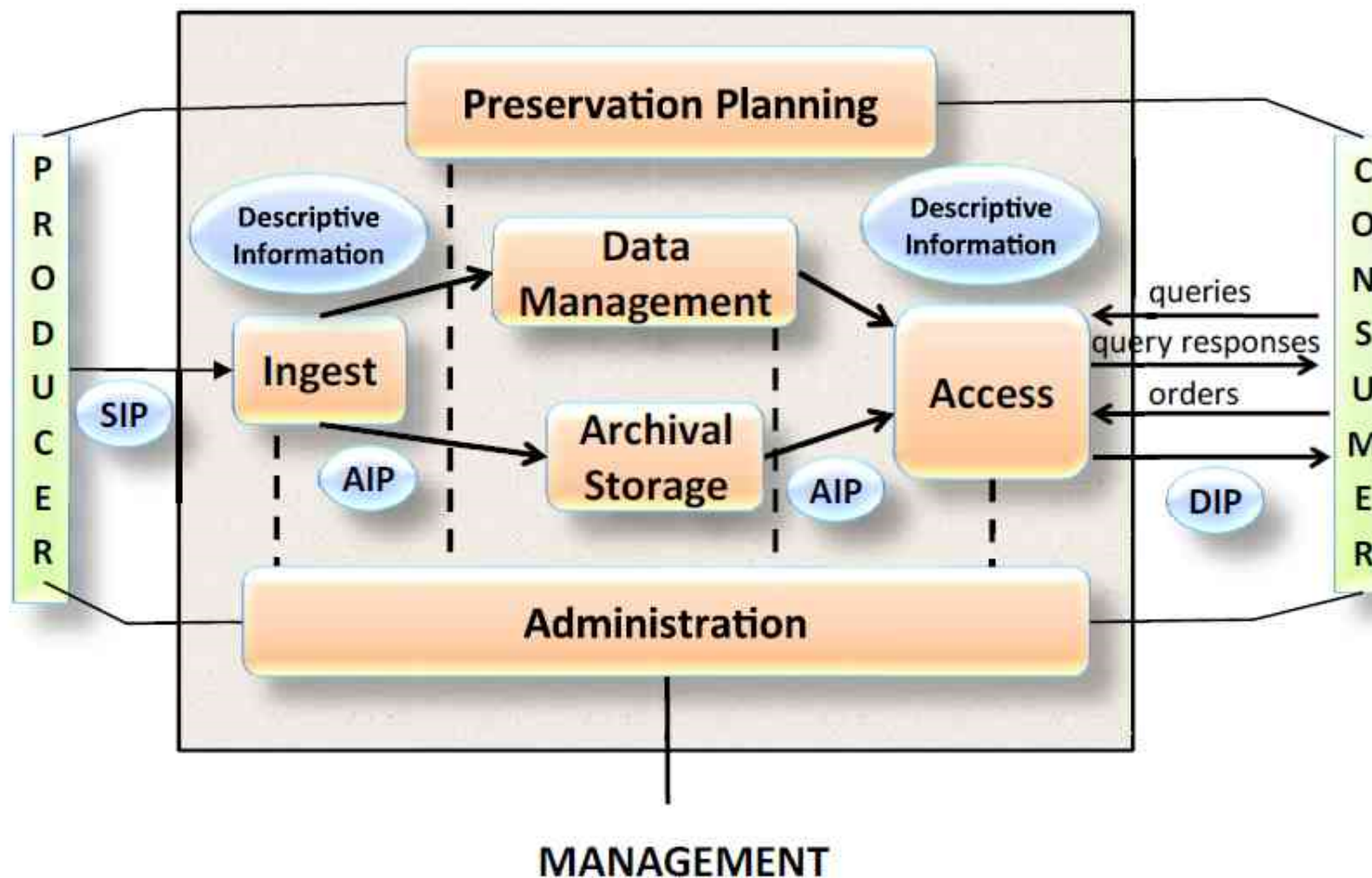
# External Data Flow View



# Outline

- 
- Principles of the OAIS Model
  - Technical Overview
  - Functional Overview
  - Information Modell
  - Summary
-

# Open Archival Information System: Six Functional Entities



SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package



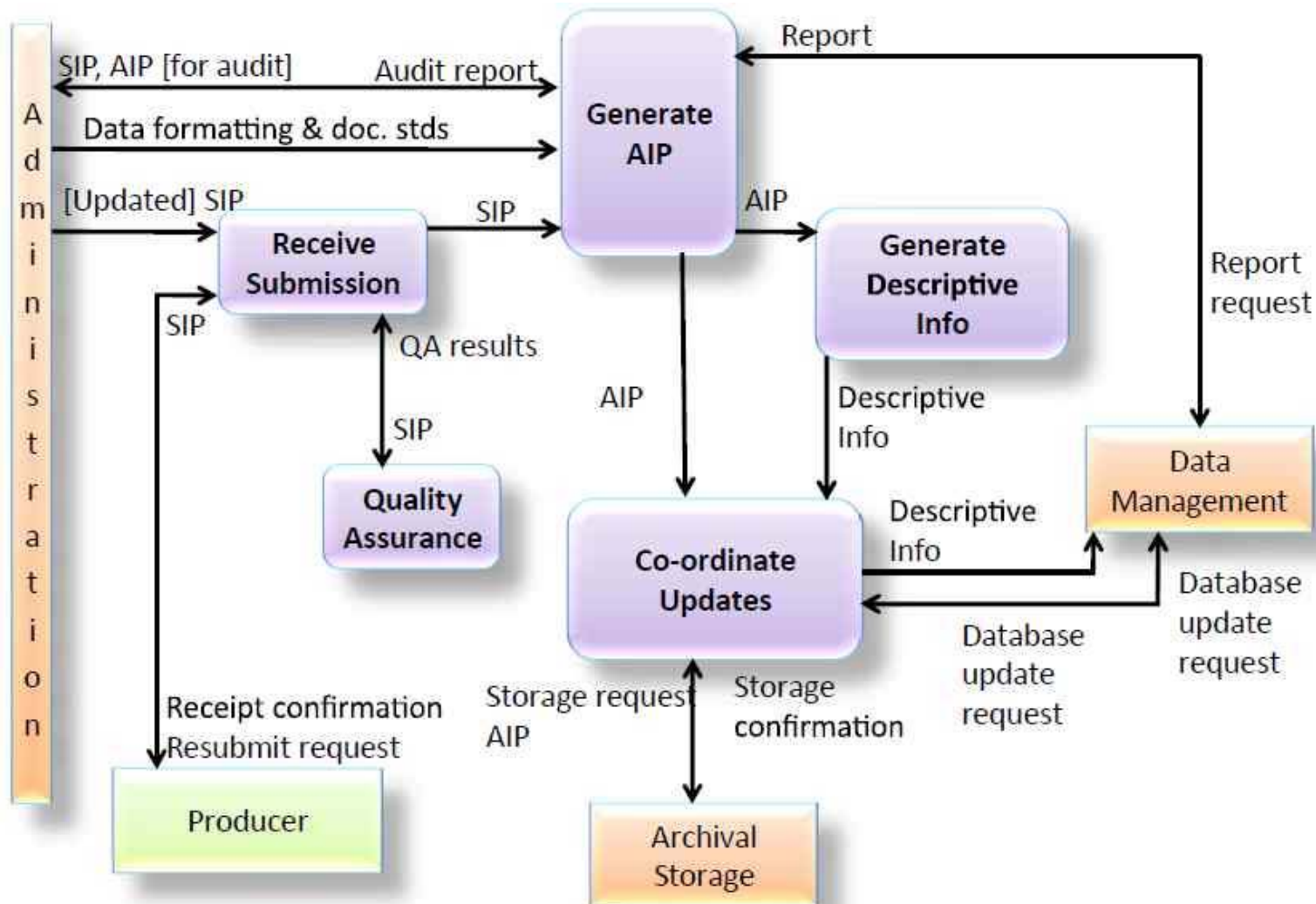
# Functional Entities in an OAIIS (1/2)

- **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.

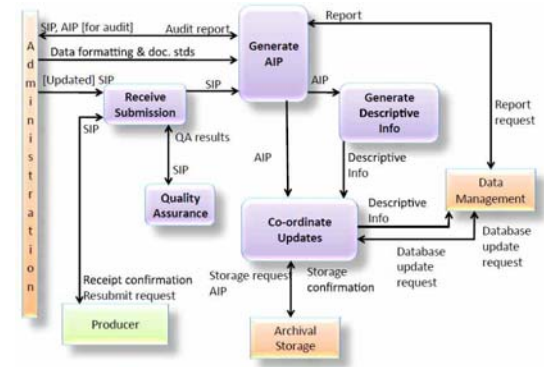
## Functional Entities in an OAIS (2/2)

- **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

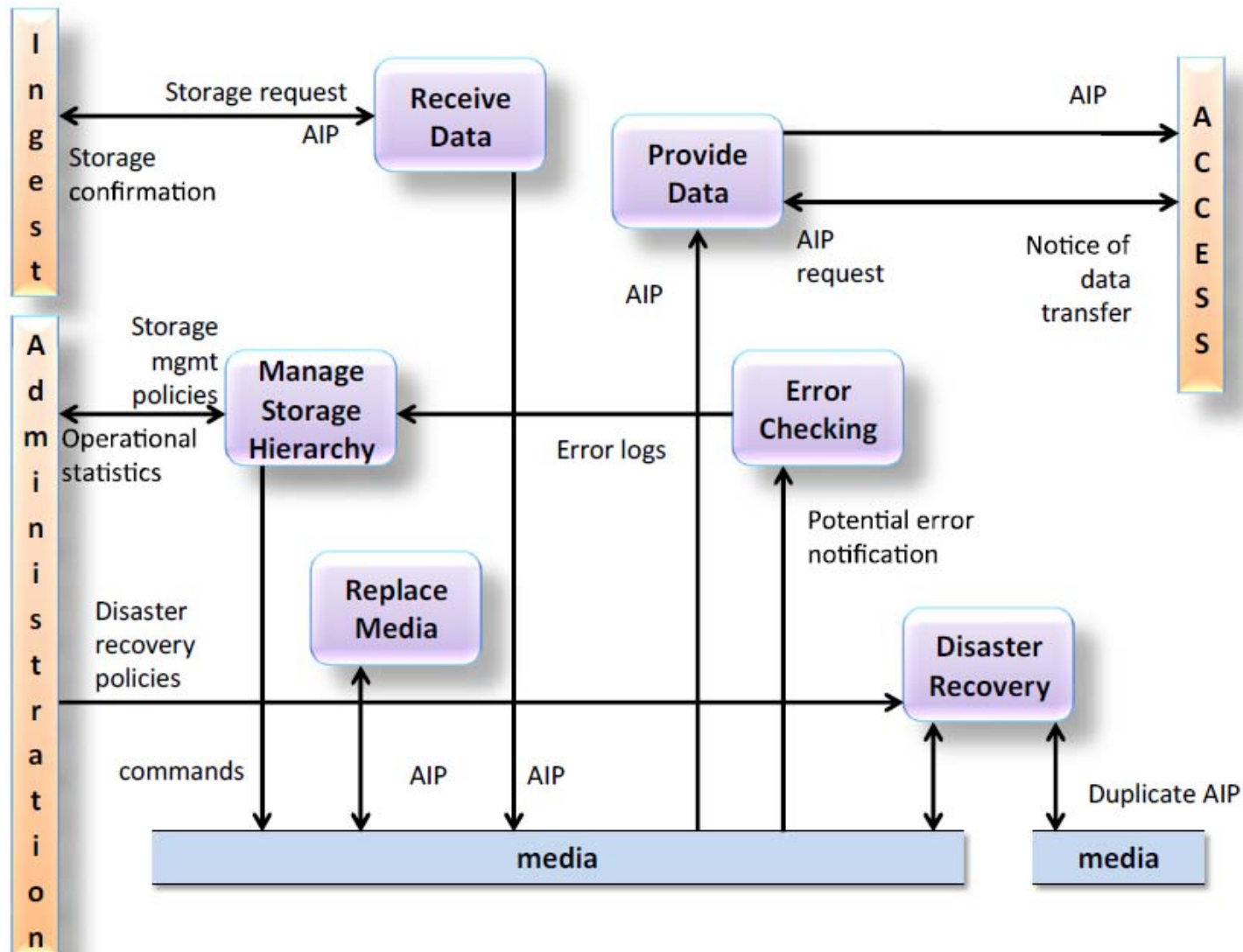
# Ingest Data Functions



# Ingest Data Functions

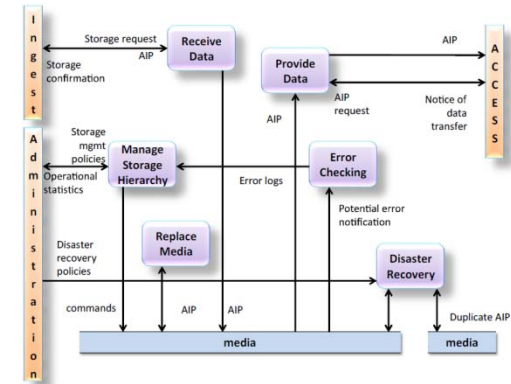


# Archival Storage Functions

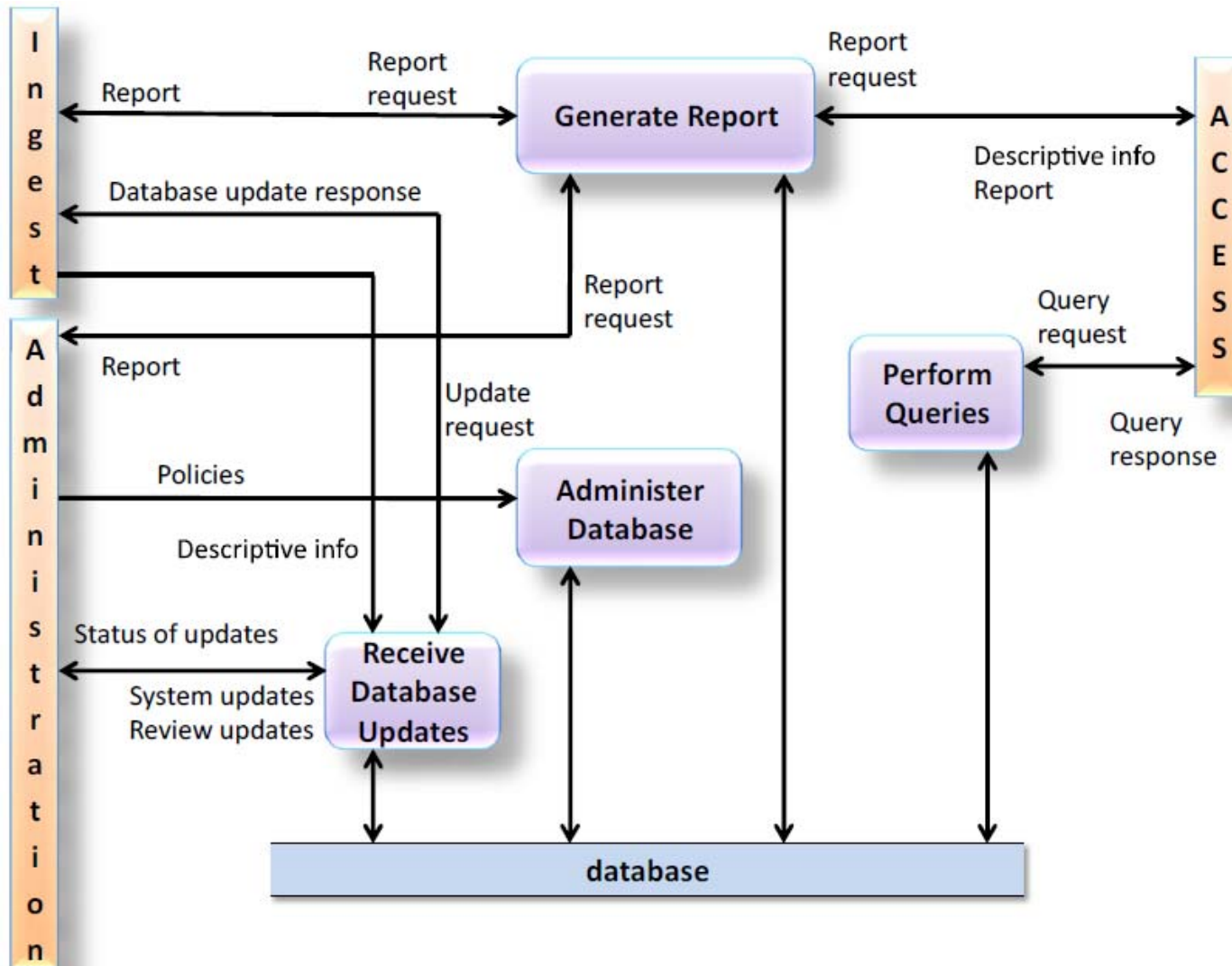


# Archival Storage Functions

- **Receive Data:**
  - accept storage request for AIP
  - Decides on storage location, media
  - Returns confirmation messages
- **Manage Storage Hierarchy**
  - Management of storage according to policies
  - Monitoring of error messages, operational statistics
- **Replace Media**
  - Reproduction of AIPs over time (no changes of content or Preservation Description Information, only Packaging Information – other changes need to go via Administration)
- **Error Checking**
  - PDI Fixity Information (CRCs, error-correcting codes, ...)
- **Disaster Recovery**
  - Duplicating of storage media content (back-up)
  - Transport to physically separated location
- **Provide Data**
  - Generate copies of AIPs for Access

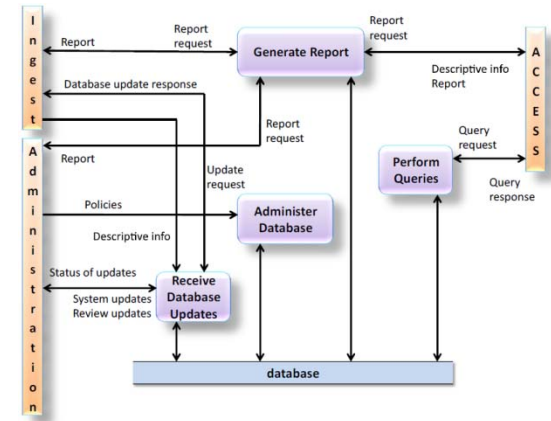


# Data Management Functions



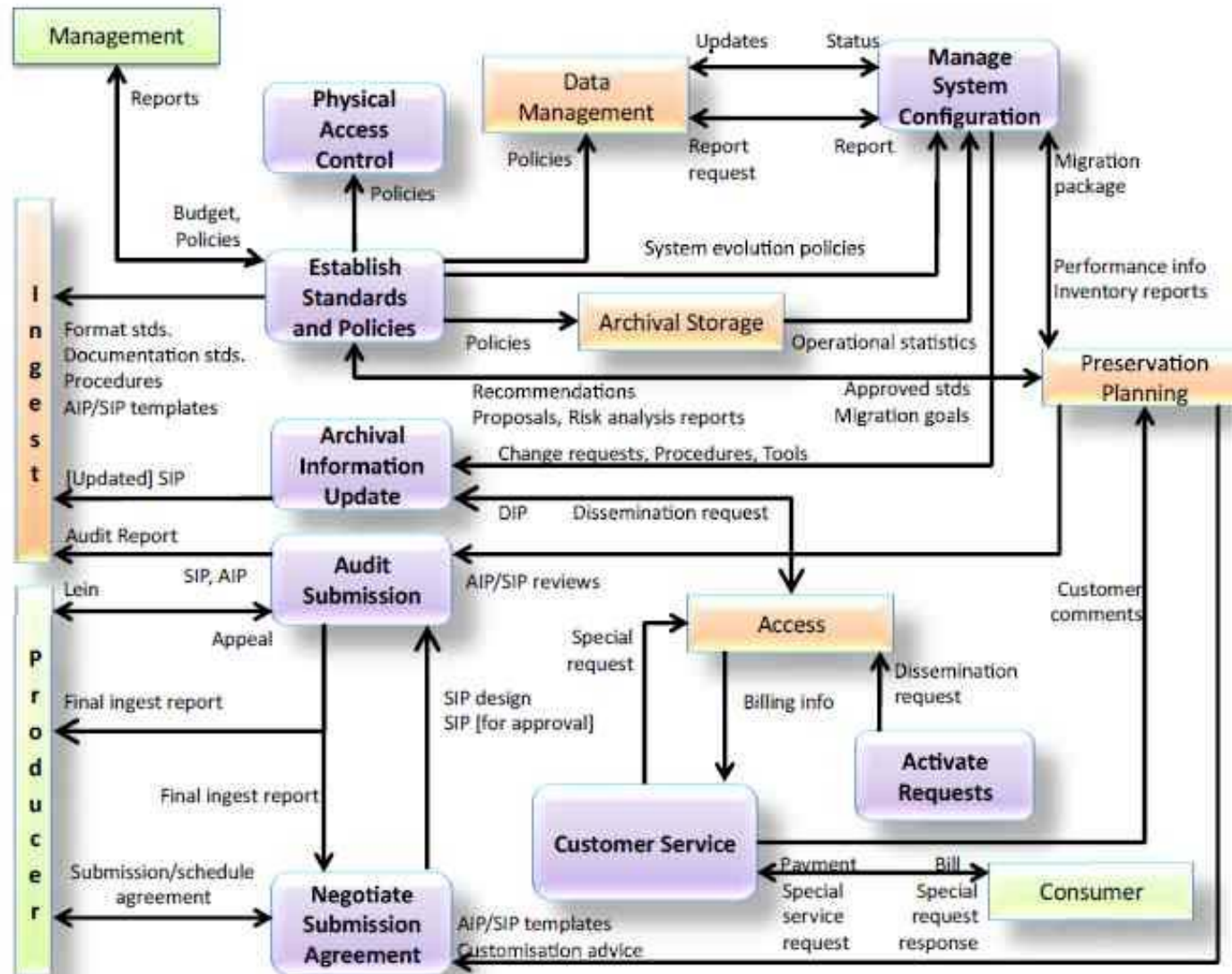
# Data Management Functions

- Administer Database
  - Integrity of DB for Descriptive Information and system information
- Perform Queries
  - Processing of queries from Access
- Generate Reports
  - Reports for Ingest, Access, Administration
- Receive DB Updates
  - Add/delete/modify information in Management DB
  - Ingest: new AIPs, Administration: updates



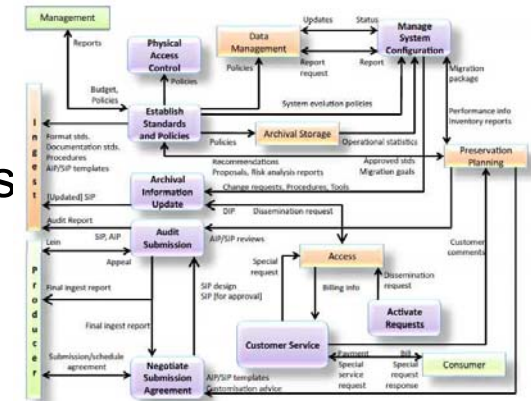


# Administration Functions



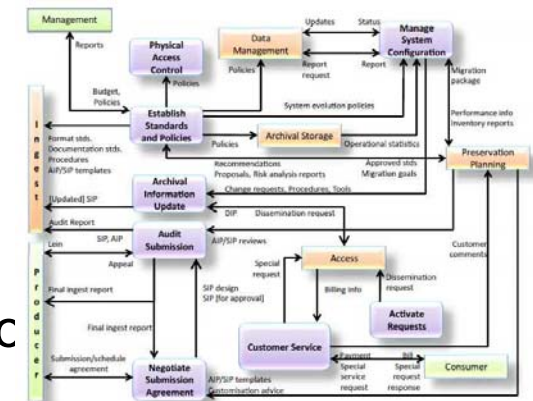
# Administration Functions

- **Negotiate Submission Agreement**
  - Contracts with Producers, submission procedures
- **Manage System Configuration**
  - System evolution, monitoring
  - Provide information for Policies
- **Archival Information Update**
  - Update content of the archive: Modifying DIPs and Re-Submission -> Migration
- **Establish Standards and Policies**
  - Budget, Standards, Policies
- **Audit Submission**
  - Analyse whether SIPs and AIPs conform to policies and regulations
  - Verifia Representation and Package Information

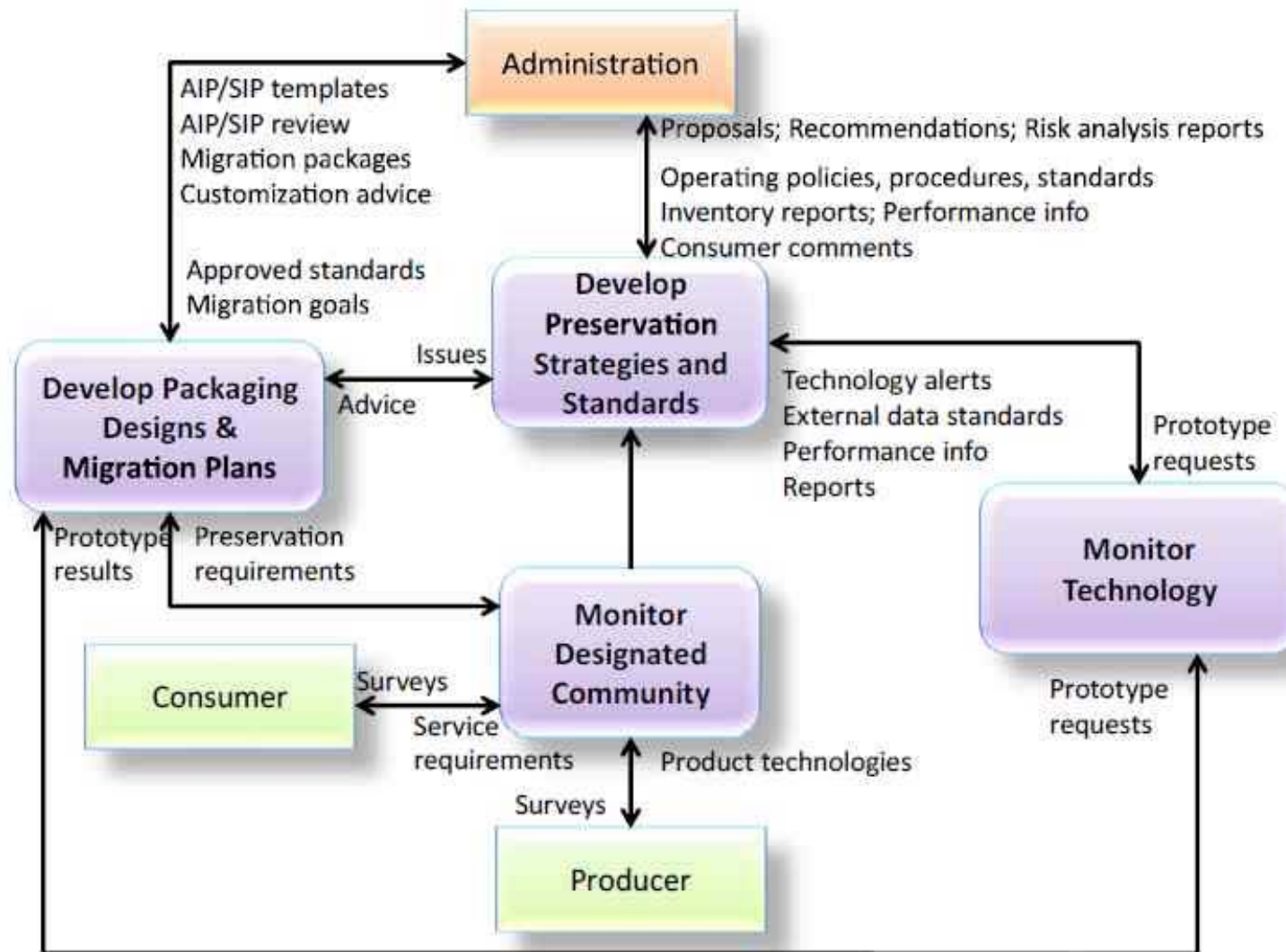


# Administration Functions

- **Activate Requests**
  - Protocol of triggered queries/evaluations
  - Periodic checks/queries to archive to verify cc
  - Ordering data/reports periodically
- **Customer Service**
  - Manage customer accounts
  - Collect costs from Access, create invoices for customers

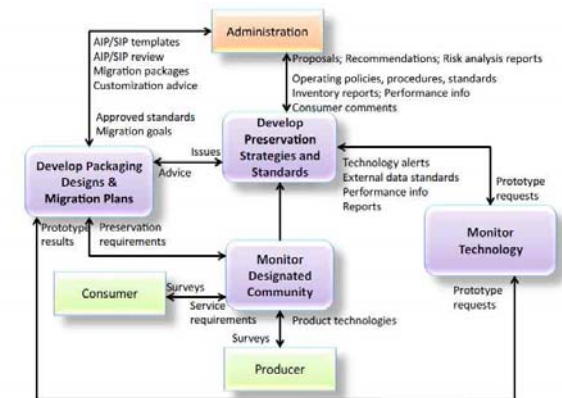


# Preservation Planning Functions

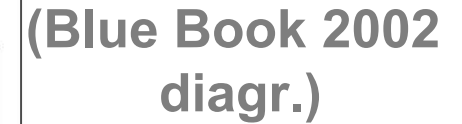


# Preservation Planning Functions

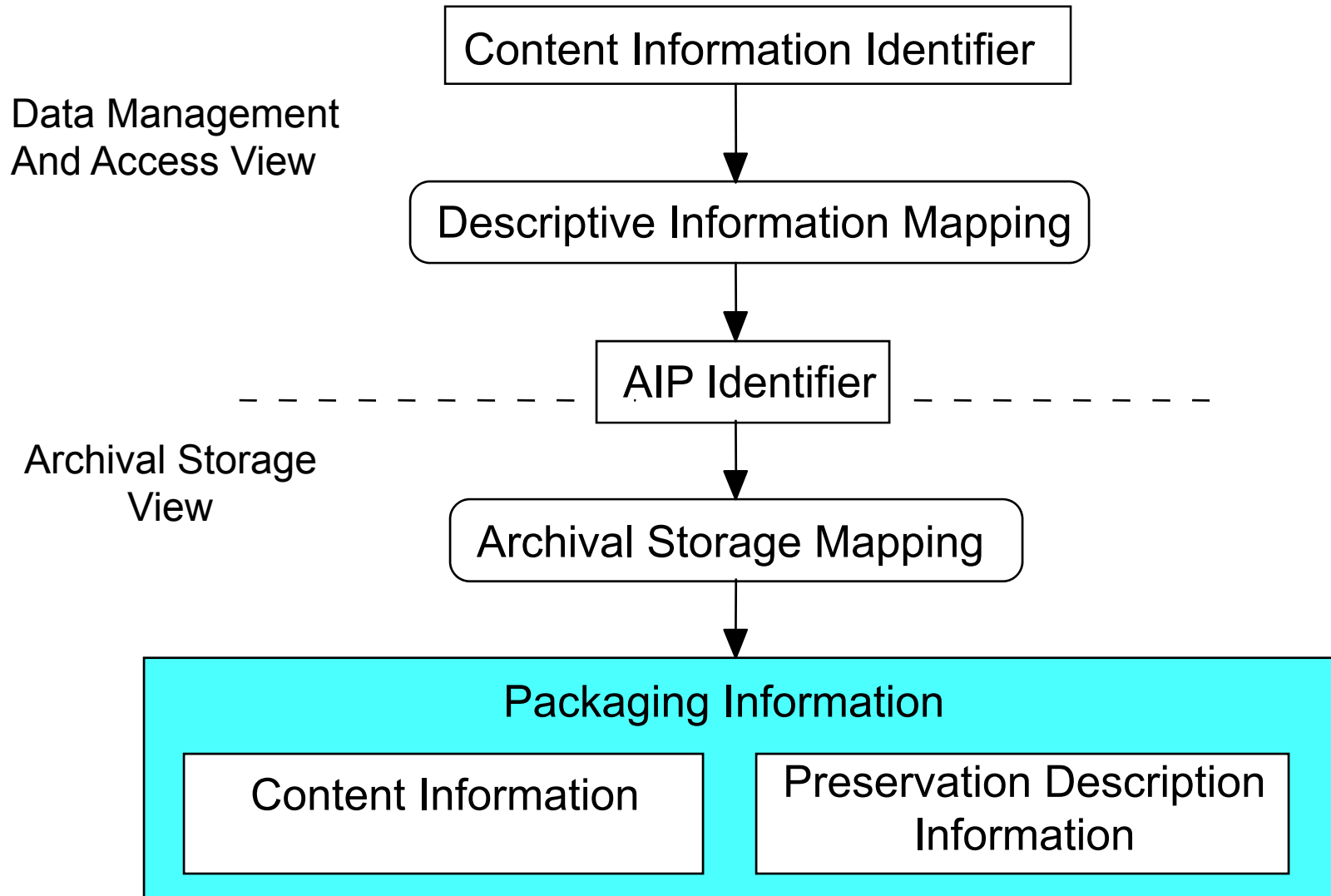
- Monitor Designated Community
  - Interaction with Producer and Consumer
- Monitor Technology
  - Technology evolution: HW, SW, Formats
- Develop Preservation Strategies and Standards
  - Strategies, trend analysis
- Develop Packaging Designs and Migration Plans
  - Migration paths, tools
  - Create Preservation Description Information







# Migration Context



# Digital Migration

Digital Migration is defined to be the transfer of digital information, while intending to preserve it, within the OAIS.

- Focus on preservation of the full information content
- New information implementation replaces the old
- OAIS has full control and responsibility over all aspects of the transfer



# Migration Motivators

- Motivators driving digital migrations
  - Media Decay
    - Often this is superceded by escalating media drive maintenance costs
  - Increased Cost Effectiveness
    - More cost-effective media types with higher volumes and lower drive maintenance costs
  - New User/Consumer Service Requirements
    - New formats more compatible with user's technology and applications
  - Proprietary software evolution
    - New software versions used to 'upgrade' formats of the information objects being preserved

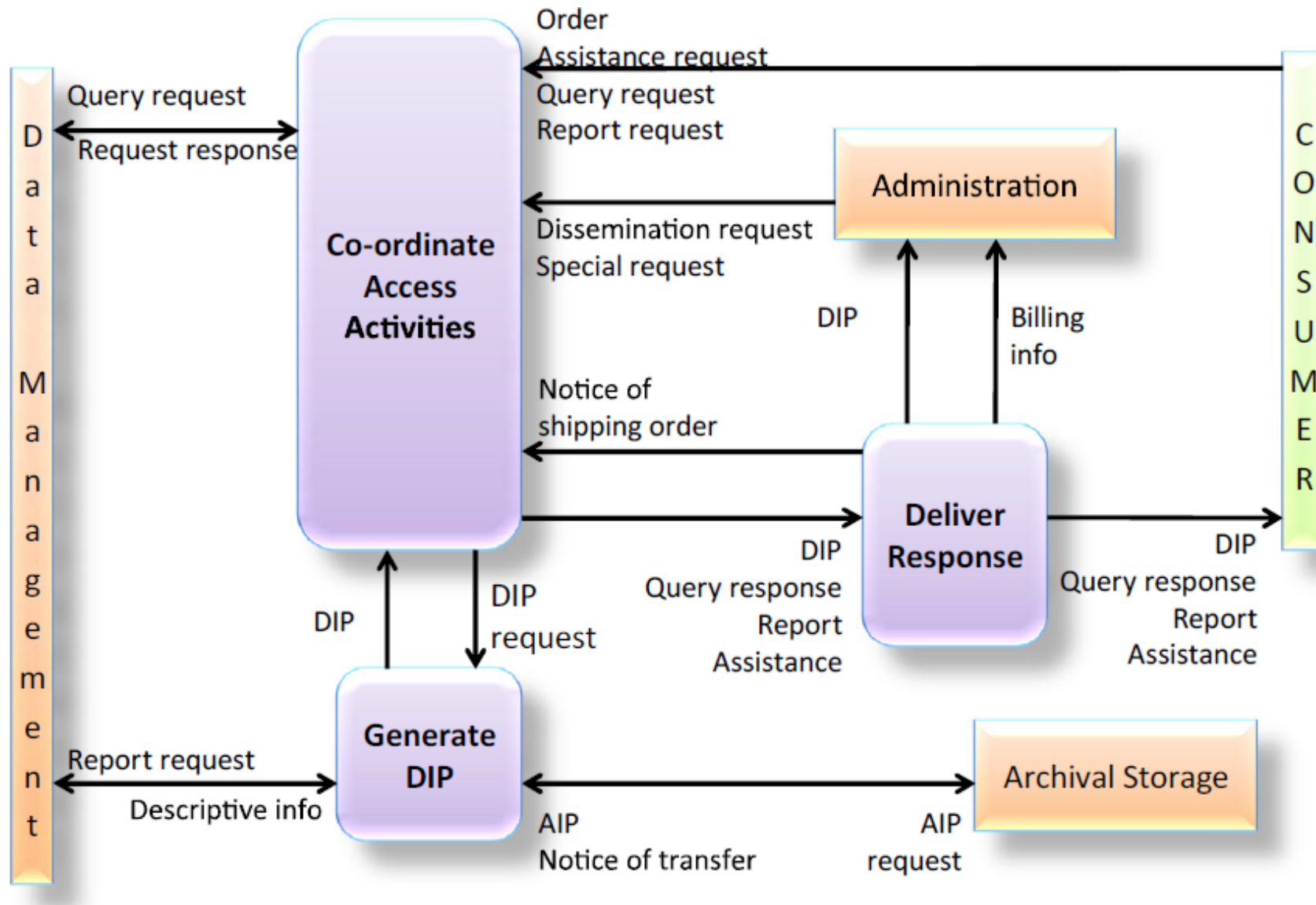
# Digital Migration Approaches

- Four primary types of digital migration in response to motivators, ordered by increasing risk of information loss:
  - Refreshment
    - Media replacement with no bit changes
  - Replication
    - No change to Packaging Information or Content Information bits (e.g. copying to new file / new location)
  - Repackaging
    - Some bit changes in Packaging Information (e.g. multiple files packaged in directory structure get copied to other carrier)
  - Transformation
    - Reversible: Bit changes in Content Information are reversible by an algorithm
    - Non-reversible: Bit changes in Content Information are not reversible by an algorithm

# Digital Migration and AIPs

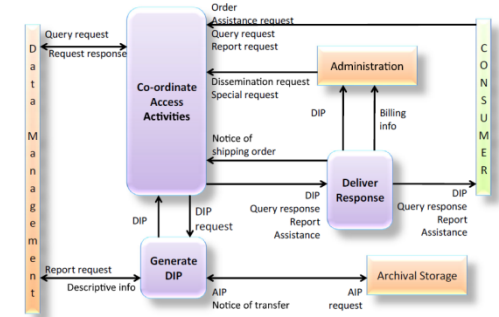
- Unless migration involves transformation:  
no new AIP version
- Transformation:  
new **AIP Version**
- Upgrading or improvement of AIPs:  
new **AIP Edition**
- Extracting or aggregating from multiple AIPs:  
**Derived AIP**

# Access Functions



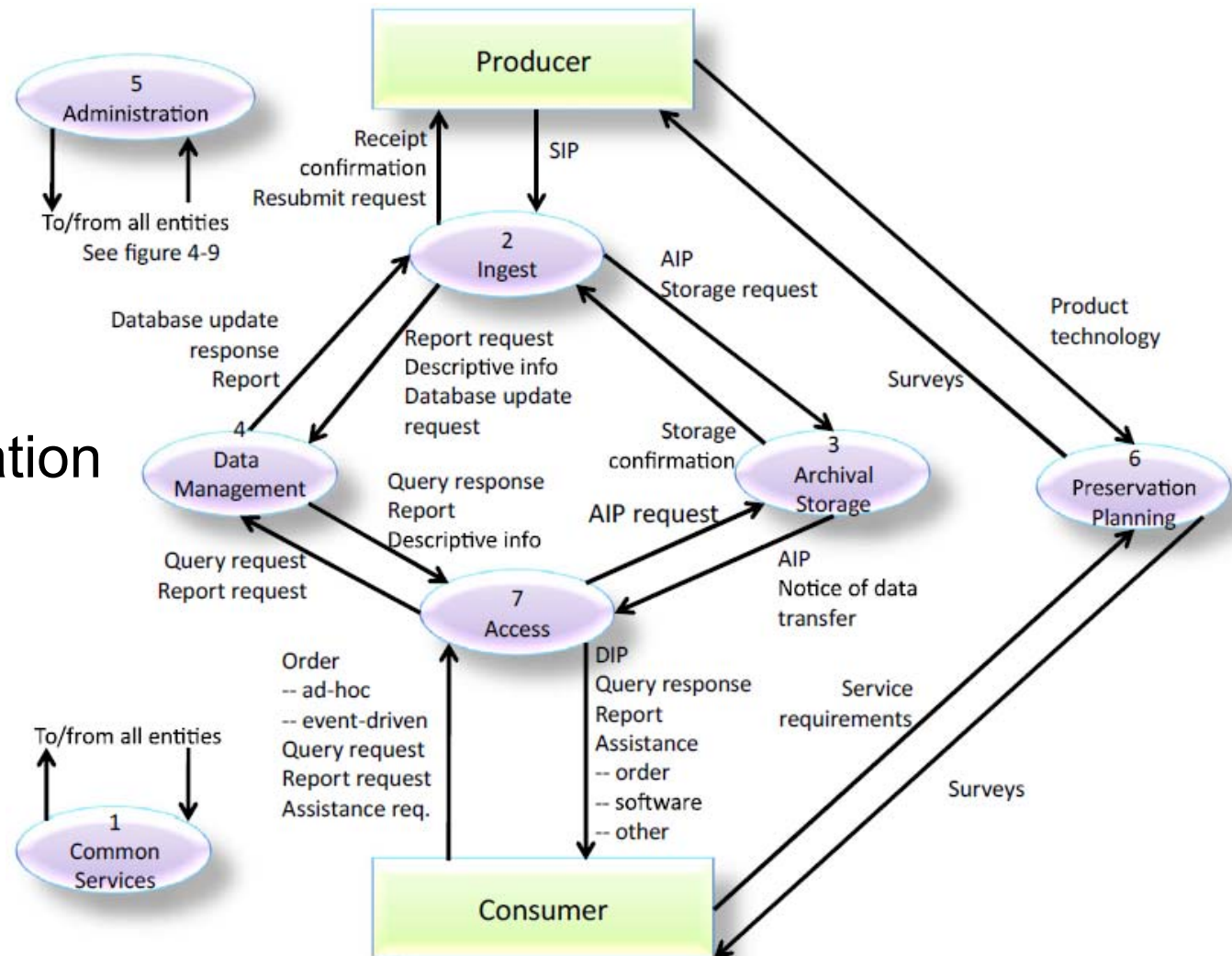
# Access Functions

- Coordinate Access Activities
  - User interface, authorization
  - 3 types of Requests:
    - Queries to Data Management for Result Set
    - Order for Data Management and Archival Storage
    - Dissemination Requests by Administration for Archival Information Update
- Generate DIP
  - Get data from Archival Storage into Staging Area
  - Get Descriptive Information from Data Management
  - Apply processes to transform AIPs into a suitable DIP depending on query / consumer
- Deliver Response
  - On-line and off-line responses
  - Forward results

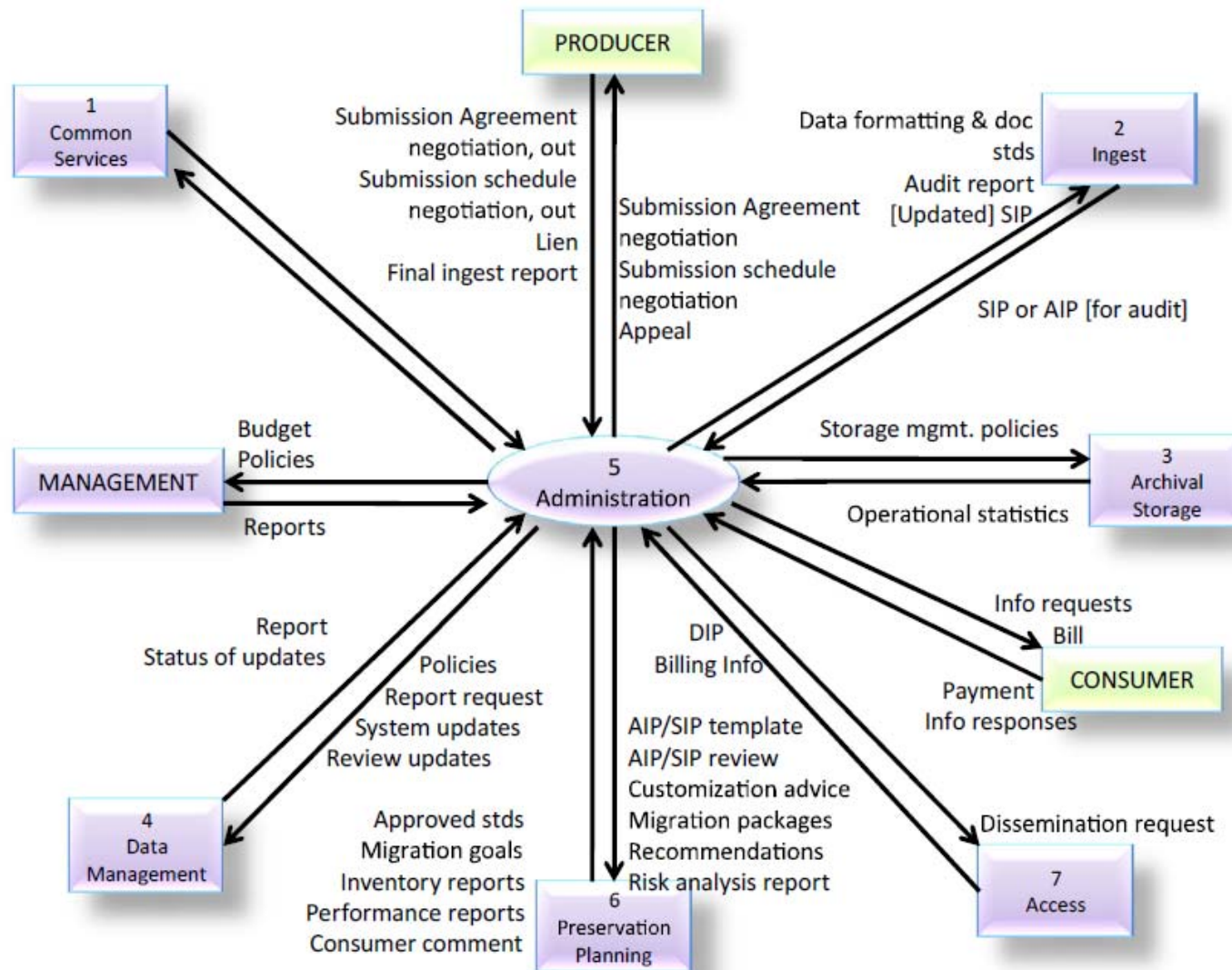


# OAIS Data Flow

■ w/o administration



# OAIS Data Flow



# Common Services

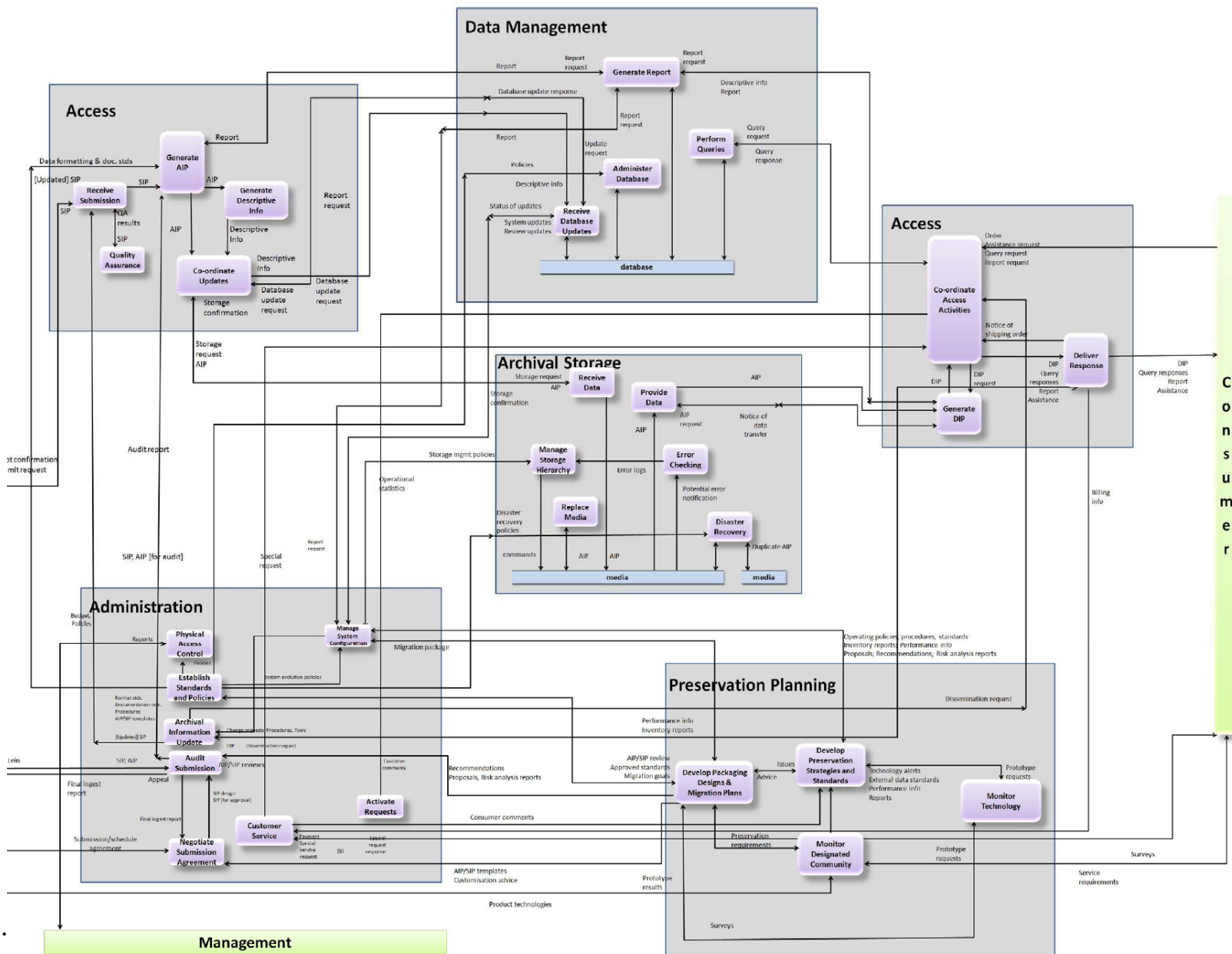
- Modern, distributed computing applications assume a number of supporting services
- Examples of Common Services include:
  - inter-process communication
  - name services
  - temporary storage allocation
  - exception handling
  - security
  - file and directory services



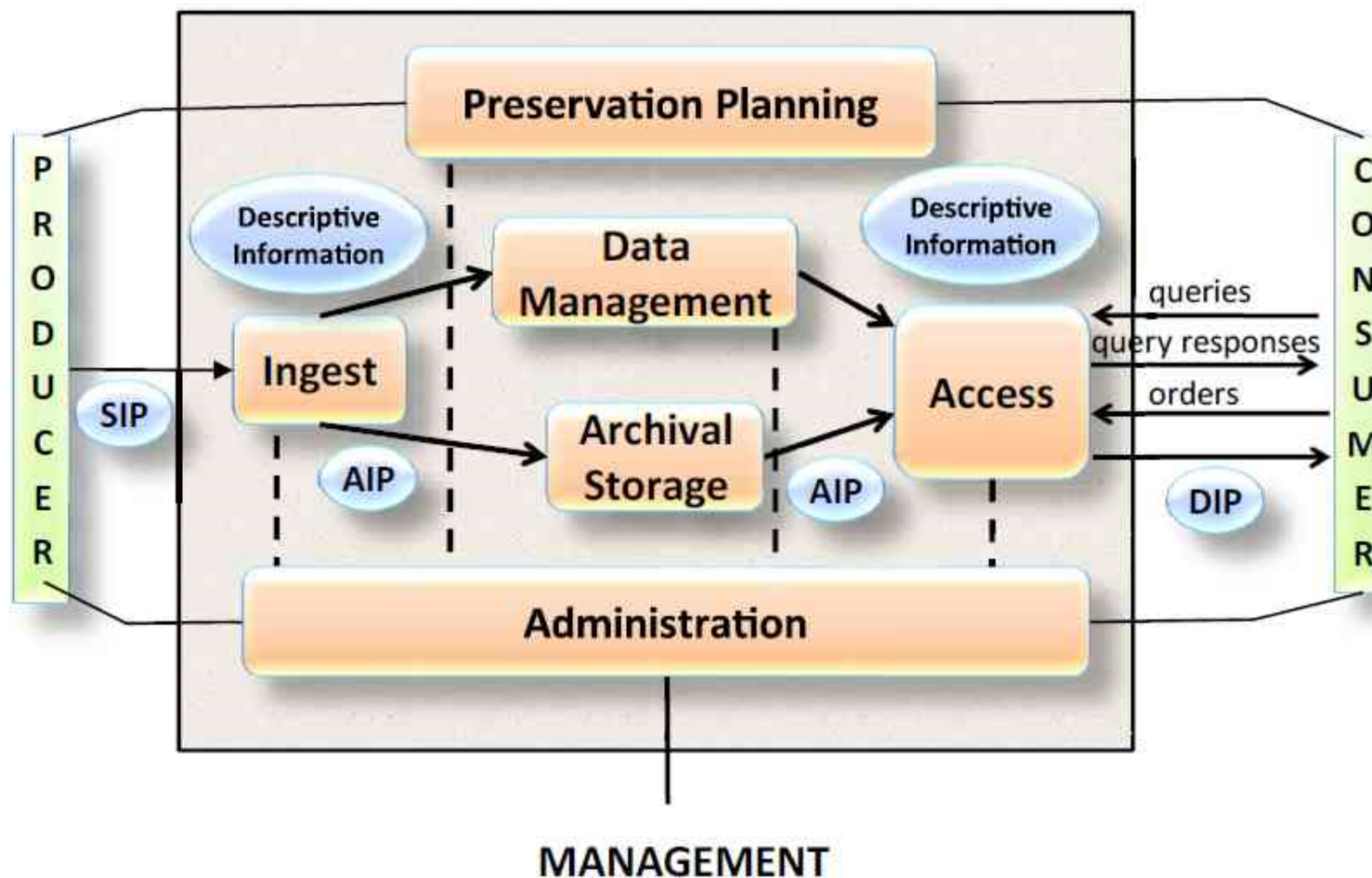
# Common Services

- Important:
  - All processing steps taken need to be documented (logs, protocols)
  - Reporting
  - Confirmations
  
- This documentation is part of the archive as well

# OAIS Composite Functional Entities



# Open Archival Information System: Summary



SIP = Submission Information Package

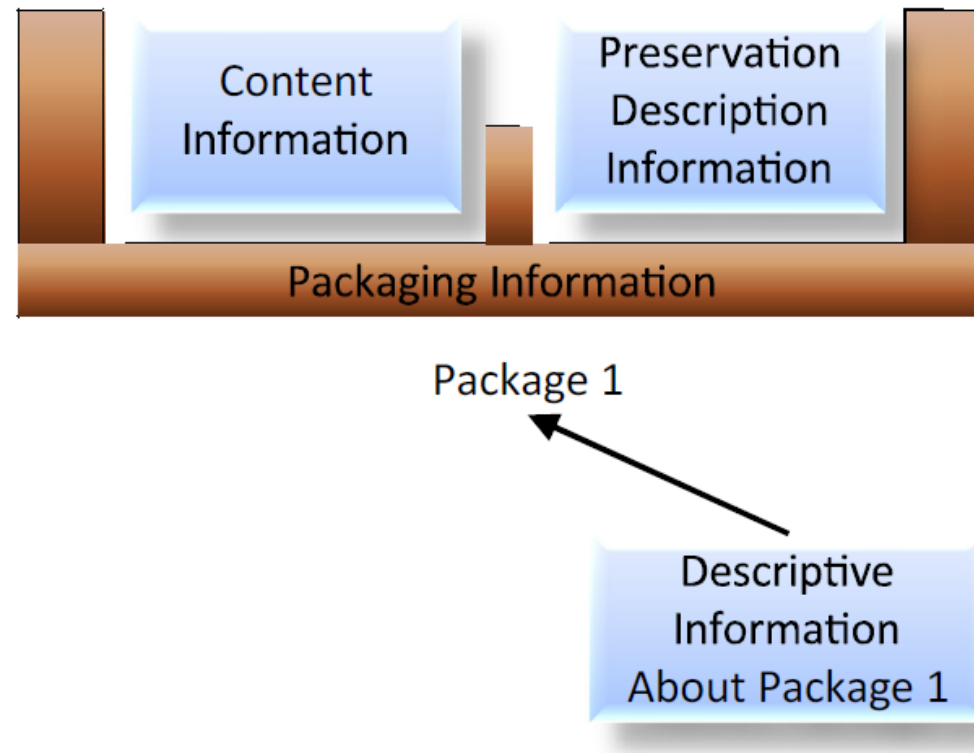
AIP = Archival Information Package

DIP = Dissemination Information Package

# Outline

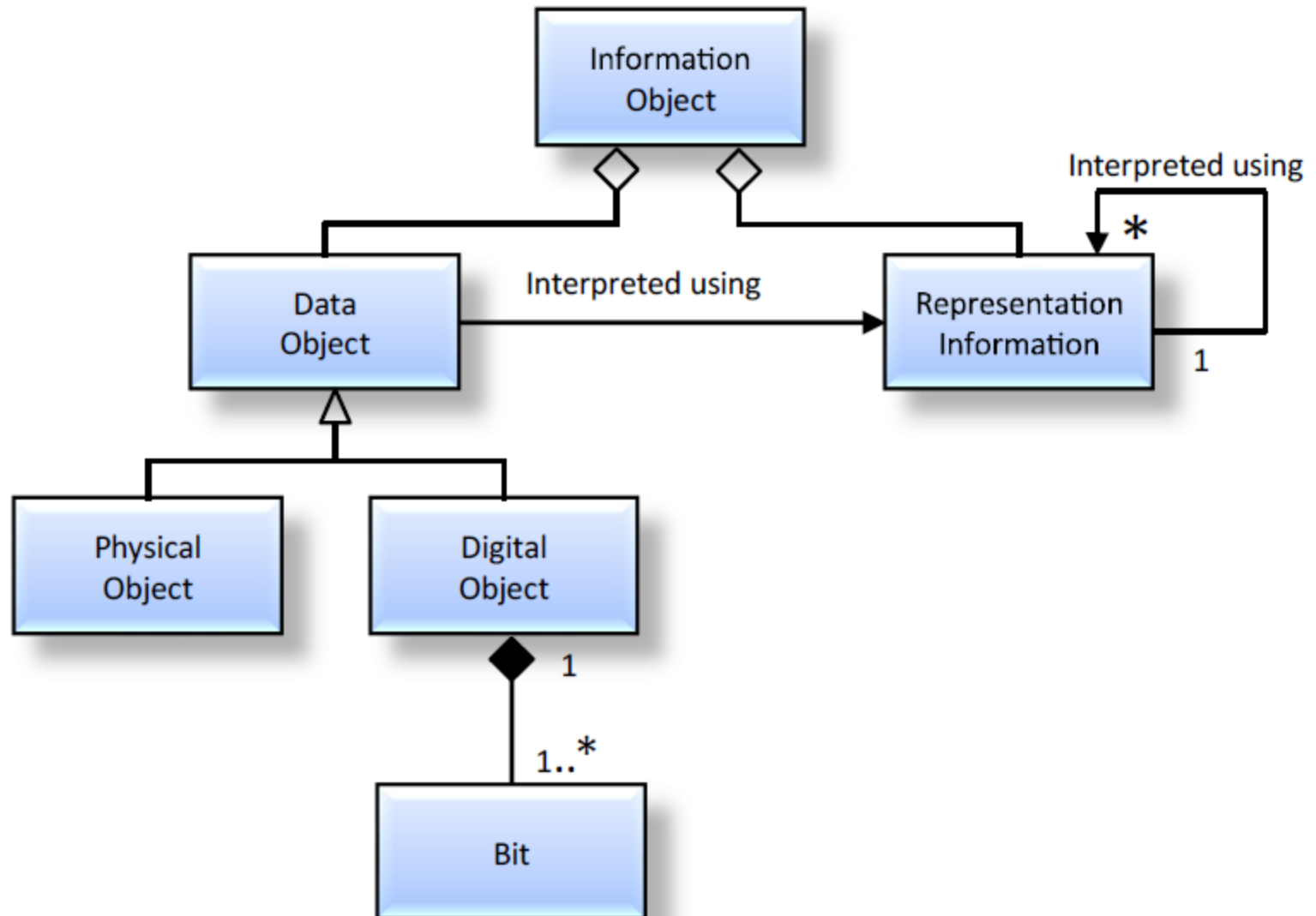
- 
- Principles of the OAIS Model
  - Technical Overview
  - Functional Overview
  - Information Modell
  - Summary
-

# Information Package Definition



- An Information Package is a conceptual container holding two types of information
  - Content Information
  - Preservation Description Information (PDI)
- Plus descriptive information

# Information Object

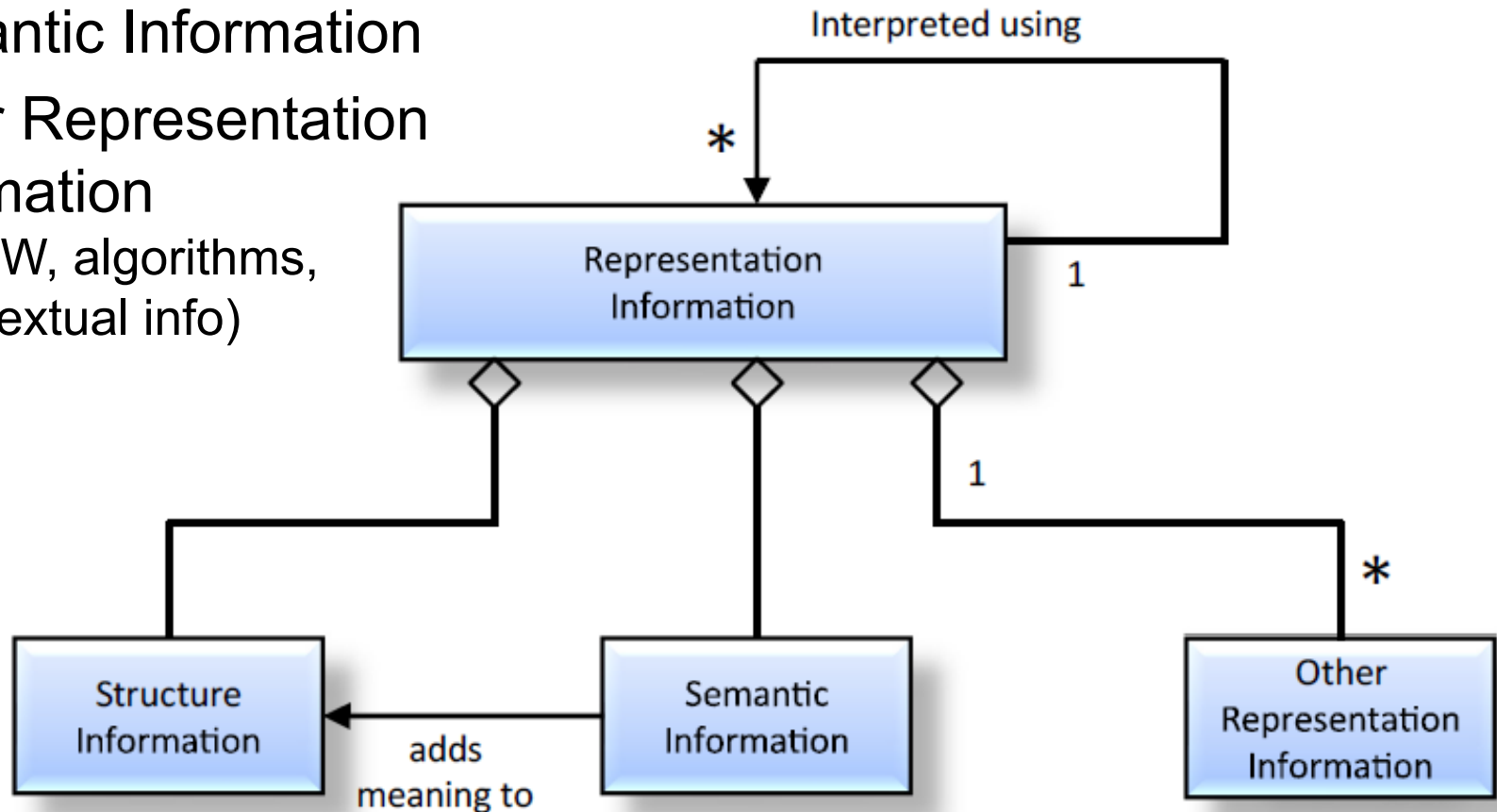


# Representation Information

- The Representation Information accompanying a physical object, like a moon rock, may give additional meaning
  - It typically is a result of some analysis of the physically observable attributes of the rock
- The Representation Information accompanying a digital object, or sequence of bits, is used to provide additional meaning.
  - It typically maps the bits into commonly recognized data types such as character, integer, and real and into groups of these data types.
  - It associates these with higher level meanings which can have complex inter-relationships that are also described

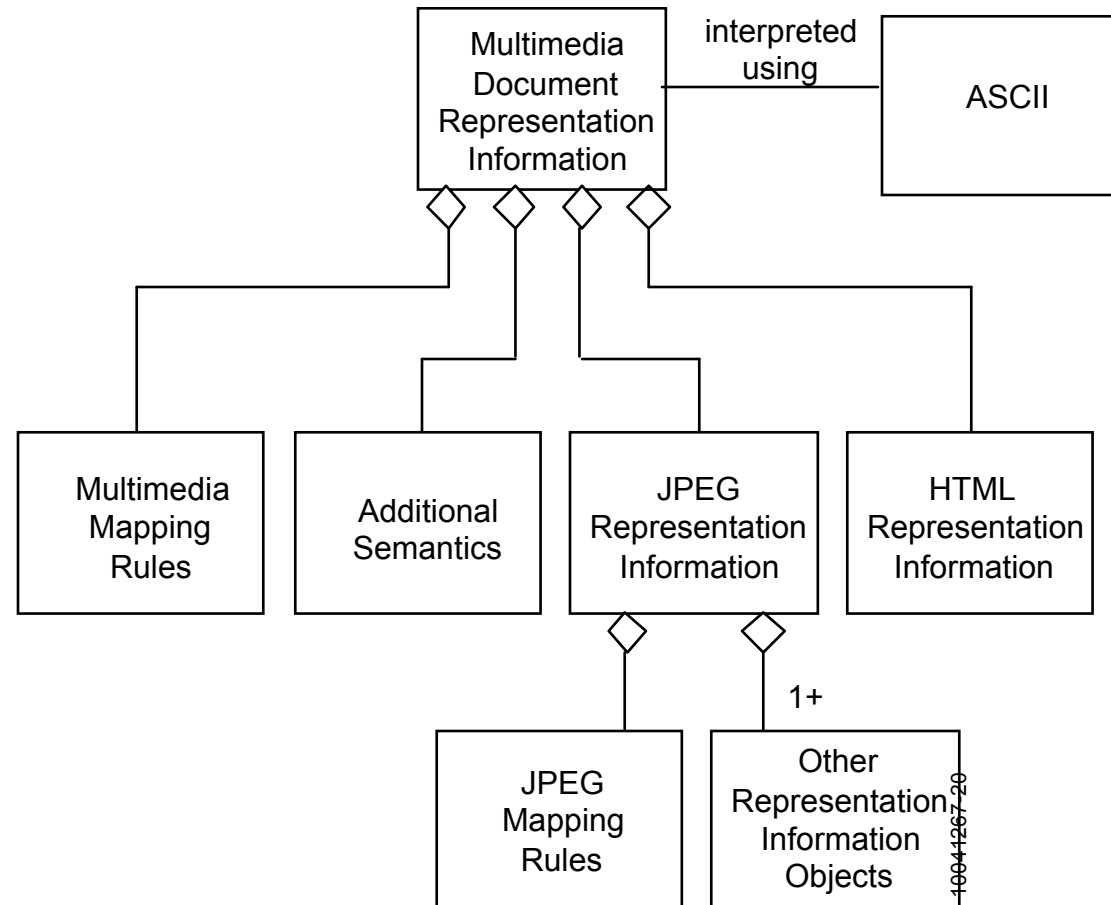
# Recursive Nature of Representation Information

- Structure Information
- Semantic Information
- Other Representation Information  
(e.g. SW, algorithms, other textual info)

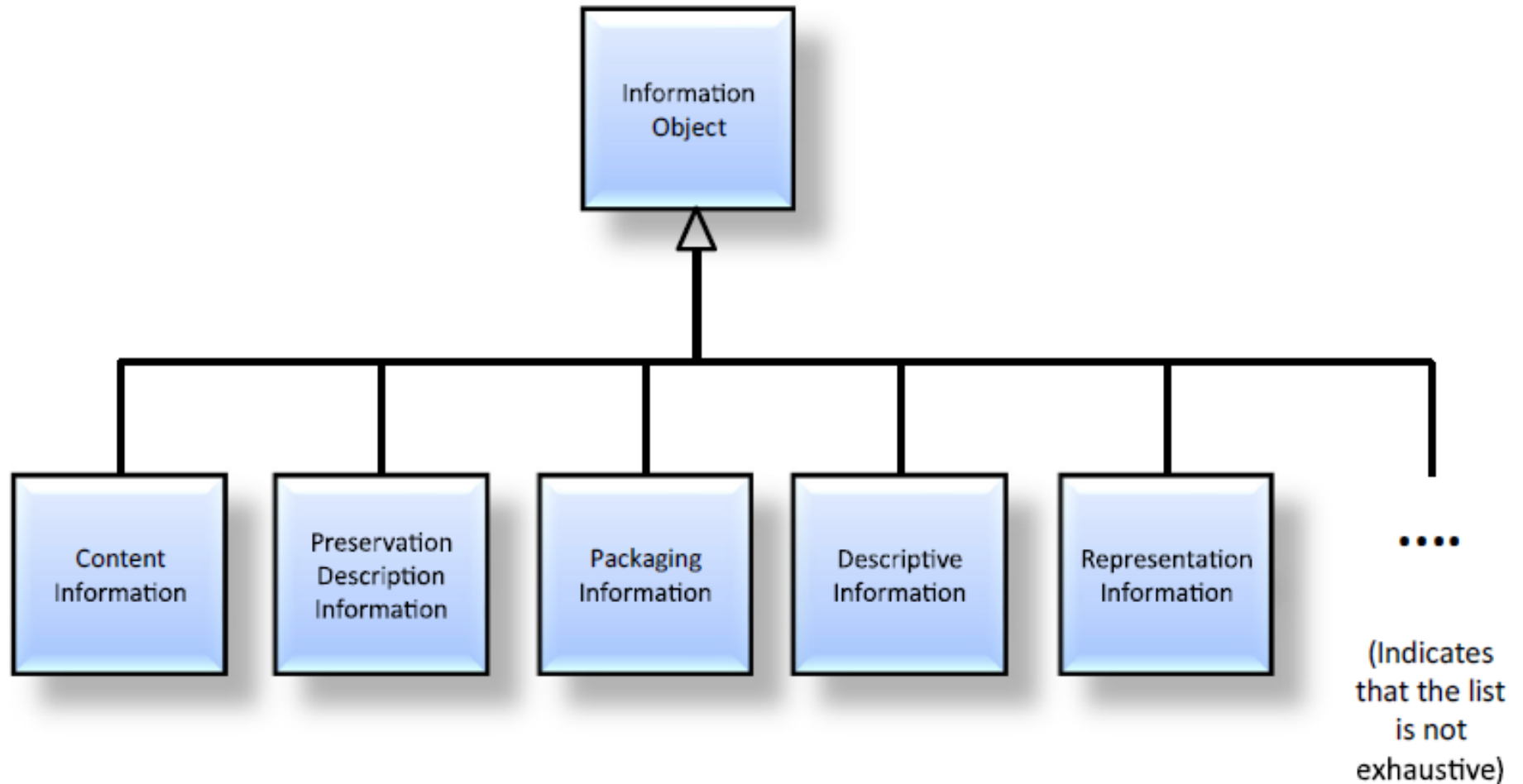




# Sample Representation Net



# Types of Information Used in OAIS



# Content Information

- The information which is the primary object of preservation
- An instance of Content Information is the information that an archive is tasked to preserve.
- Deciding what is the Content Information may not be obvious and may need to be negotiated with the Producer
- The Data Object in the Content Information may be either a Digital Object or a Physical Object (e.g., a physical sample, microfilm)

- **Provenance Information**

- Describes the source of Content Information, who has had custody of it, what is its history

- **Context Information**

- Describes how the Content Information relates to other information outside the Information Package

- **Reference Information**

- Provides one or more identifiers, or systems of identifiers, by which the Content Information may be uniquely identified

- **Fixity Information**

- Protects the Content Information from undocumented alteration

# PDI Examples

Content Information Type	Reference	Provenance	Context	Fixity	Access Rights
Space Science Data	<ul style="list-style-type: none"> <li>Object identifier</li> <li>Journal reference</li> <li>Mission, instrument, title, attribute set</li> </ul>	<ul style="list-style-type: none"> <li>Instrument description</li> <li>Principal Investigator</li> <li>Processing history</li> <li>Storage and handling history</li> <li>Sensor description</li> <li>Instrument</li> <li>Instrument mode</li> <li>Decommutation map</li> <li>Software interface specification</li> <li>Information Property Description</li> </ul>	<ul style="list-style-type: none"> <li>Calibration history</li> <li>Related data sets</li> <li>Mission</li> <li>Funding history</li> </ul>	<ul style="list-style-type: none"> <li>CRC</li> <li>Checksum</li> <li>Reed-Solomon coding</li> </ul>	<ul style="list-style-type: none"> <li>Identification of the properly authorized Designated Community (Access Control)</li> <li>Permission grants for preservation and for distribution</li> <li>Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)</li> </ul>

# PDI Examples

<b>Digital Library Collections</b>	<ul style="list-style-type: none"> <li>• Bibliographic description</li> <li>• Persistent identifier</li> </ul>	<ul style="list-style-type: none"> <li>• For scanned collections:             <ul style="list-style-type: none"> <li>• metadata about the digitization process</li> <li>• pointer to master version</li> </ul> </li> <li>• For born-digital publications:             <ul style="list-style-type: none"> <li>• pointer to the digital original</li> </ul> </li> <li>• Metadata about the preservation process:             <ul style="list-style-type: none"> <li>• pointers to earlier versions of the collection item</li> <li>• change history</li> <li>• Information Property Description</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Pointers to related documents in original environment at the time of publication</li> </ul>	<ul style="list-style-type: none"> <li>• Digital signature</li> <li>• Checksum</li> <li>• Authenticity indicator</li> </ul>	<ul style="list-style-type: none"> <li>• Legal framework(s)</li> <li>• Licensing offers</li> <li>• Specifications for rights enforcement measures applied at dissemination time</li> <li>• Permission grants for preservation and for distribution</li> <li>• Information about watermarking applied at submission and preservation time</li> <li>• Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)</li> </ul>
------------------------------------	--	--	--	---	---

# PDI Examples

Content Information Type	Reference	Provenance	Context	Fixity	Access Rights
Software Package	<ul style="list-style-type: none"> <li>Name</li> <li>Author/Originator</li> <li>Version number</li> <li>Serial number</li> </ul>	<ul style="list-style-type: none"> <li>Revision history</li> <li>Registration</li> <li>Copyright</li> <li>Information Property Description</li> </ul>	<ul style="list-style-type: none"> <li>Help file</li> <li>User guide</li> <li>Related software</li> <li>Language</li> </ul>	<ul style="list-style-type: none"> <li>Certificate</li> <li>Checksum</li> <li>Encryption</li> <li>CRC</li> </ul>	<ul style="list-style-type: none"> <li>Designated Community</li> <li>Legal framework(s)</li> <li>Licensing offers</li> <li>Specifications for rights enforcement measures applied at dissemination time</li> <li>Pointers to Fixity and Provenance Information (e.g., digital signatures, and rights holders)</li> </ul>

# Descriptive Information

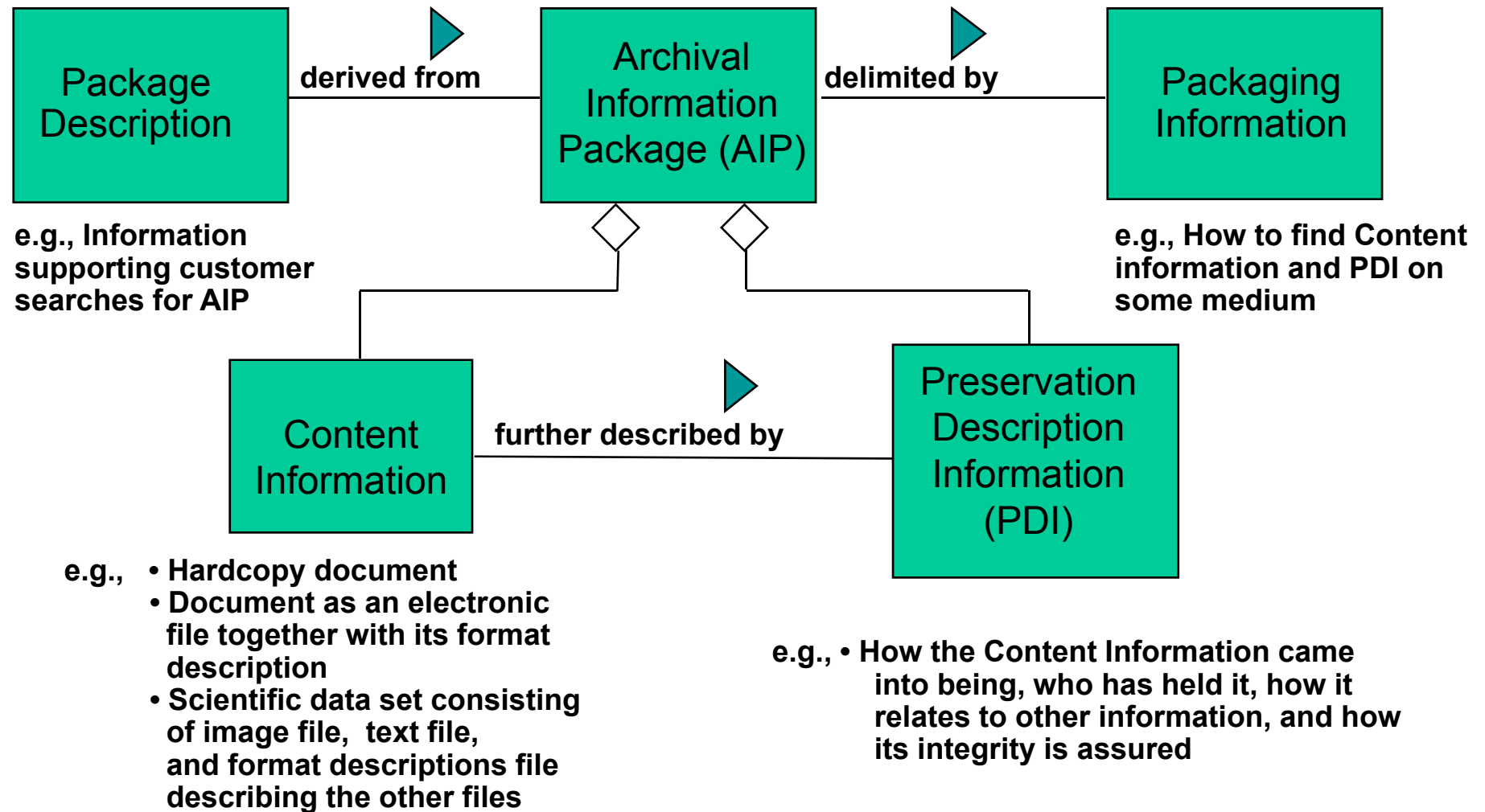
- Contain the data that serves as the input to documents or applications called Access Aids.
- Access Aids can be used by a consumer to locate, analyze, retrieve, or order information from the OAIS.



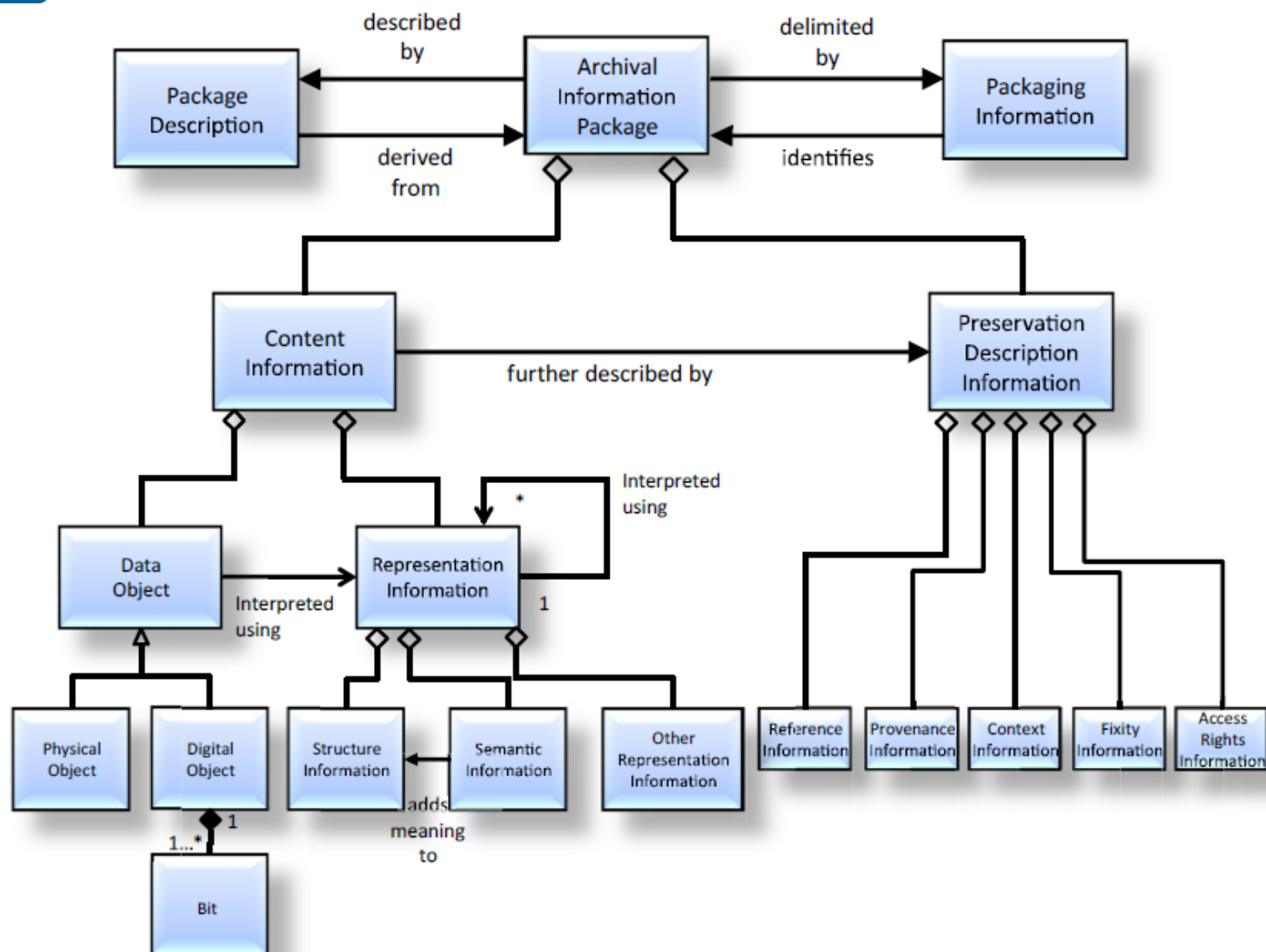
# Packaging Information

- Information which, either actually or logically, binds and relates the components of the package into an identifiable entity on specific media
- Examples of Packaging Information include tape marks, directory structures and filenames

# OAIS Archival Information Package

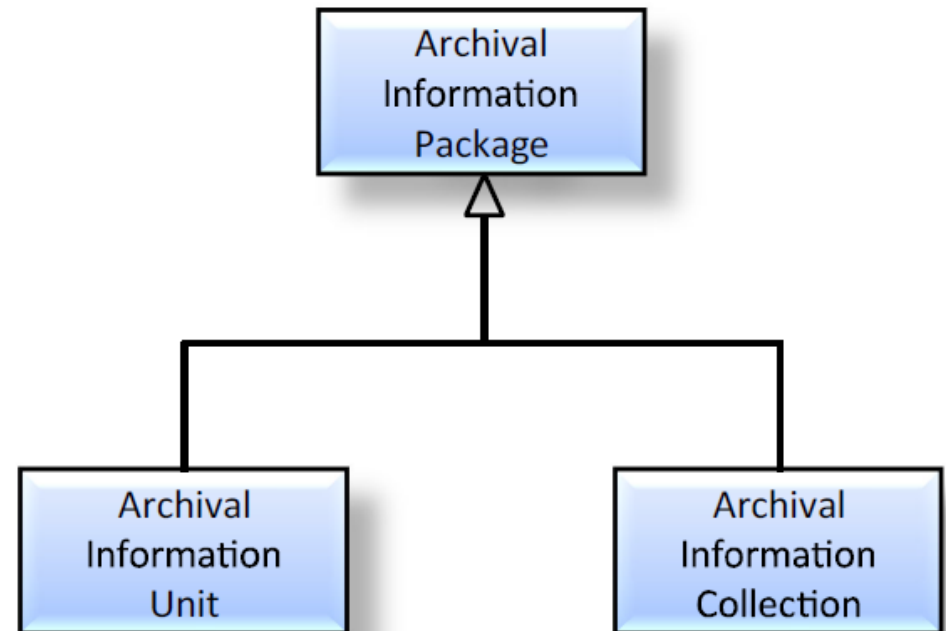


# AIP detailed view



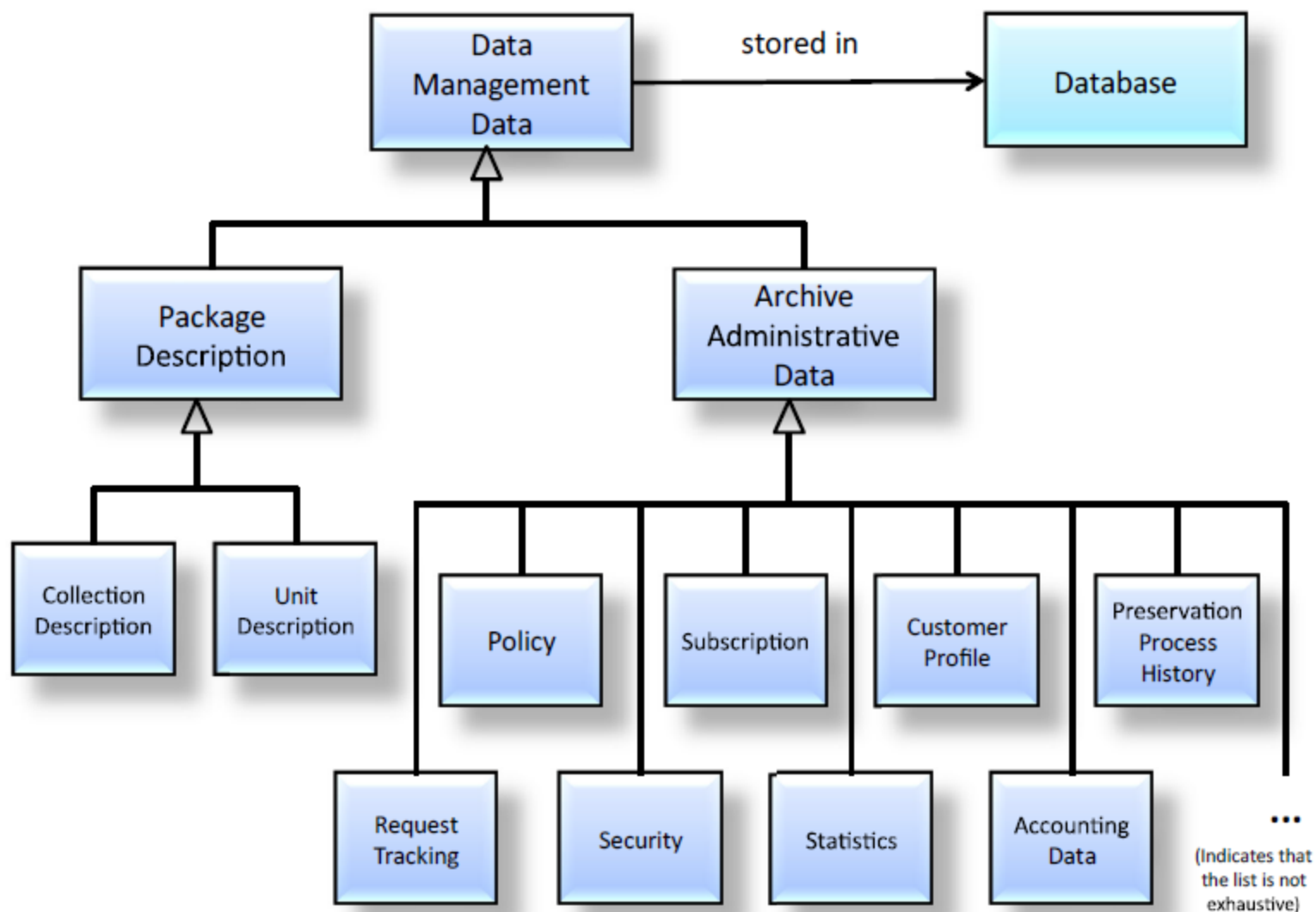
# AIP Types

- Archival Information Unit (AIU) contains a single Data Object as the Content Object
- Archival Information Collection (AIC) contains multiple AIPs in its Content Object
  - Each member of an AIC is an AIP containing Content Information and PDI
  - The AIC contains unique PDI on the collection process



- Package Descriptions are needed by an OAIIS to provide visibility and access to the OAIIS holdings
- Package Descriptions contain 1 or more Associated Descriptions which describe the AIP Content Information from the point of view of a single Access Aid
- Some example of Access Aids Include:
  - Finding Aids - assist the consumer in locating information of interest
  - Ordering Aids - allow the consumer to discover the cost of and order AIUs of interest
  - Retrieval Aids - enable authorized users to retrieve the AIU described by the Unit Descriptor from Archival Storage

# Data Management Information



# Information Model Summary

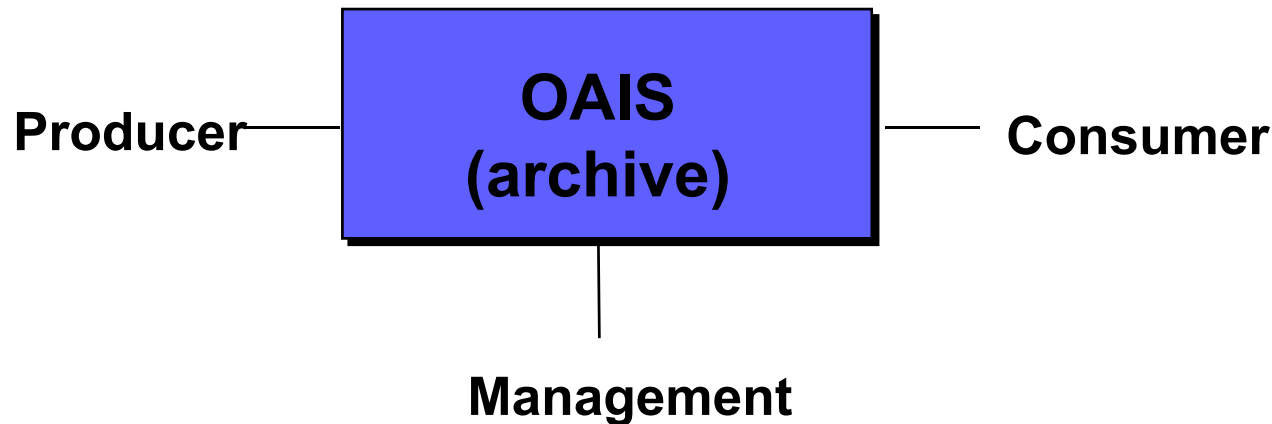
- Presented a model of information objects as containing data objects and representation objects
- Classified information required for Long-term archiving into 4 classes: Content Information, PDI, Packaging Information and Descriptive Information
- Described how these classes would be aggregated and related in an AIP to fully describe an instance of Content Information
- Presented information needed for Access, in addition to that needed for Long-term Preservation
- Put the Access oriented structures in the context of the other data needed to operate an OAIS

# Outline

- 
- Principles of the OAIS Model
  - Technical Overview
  - Functional Overview
  - Information Modell
  - Summary
-



# 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

# OAIS Information Definition

- Information is always expressed (i.e., represented) by some type of data
- Data interpreted using its Representation Information yields Information
- Information Object preservation requires clear identification and understanding of the Data Object and its associated Representation Information



# Summary

- OAIS is a reference model
- OAIS no implementation specification
- Defines language, responsibilities, functionalities,...
- Can be used for all kind of archives, institutions, organizations, systems
- Can be used for all kinds of objects, physical or digital