



*Digital Preservation:
Current Thinking*

Anne Gilliland-Swetland
Department of Information Studies



Risks to Your Investment in Digital Collections

- # Technological obsolescence
 - # Accidental or deliberate alteration resulting in diminution or loss of physical and/or intellectual integrity
 - # Limitations on preservation choices imposed by copyright legislation and licensing agreements
 - # Lack of economic models and metrics for long-term preservation
 - # Lack of sustained and sustainable preservation management infrastructure – fiscal and human resources and technological tools
-

What's New?

- # Focus on automating as many preservation processes as possible, e.g., ingestion, migration, metadata creation
 - # Not trying to preserve originals – trying to continue to preserve the ability to reproduce authentic copies of the original
 - # Metadata will have to be subject to a preservation management regime too
-

What is the Same?

- # Ideal choices for preservation may be circumscribed by and need to be managed within an economic and risk management framework established with all your collections in mind
 - # The motivation and strategy for preservation, and the object's physical and intellectual needs must be aligned
 - # A single preservation approach may not fit the needs of all your information objects
-

How These Risks Are Currently Being Addressed

- # Emerging technologies and technological approaches
 - XML, e.g., Persistent Archives and Archivists' Workbench
 - Emulation, e.g., CAMiLEON
 - # Structural and Collaborative Models for Preservation Repositories
 - # Preservation Metadata
-

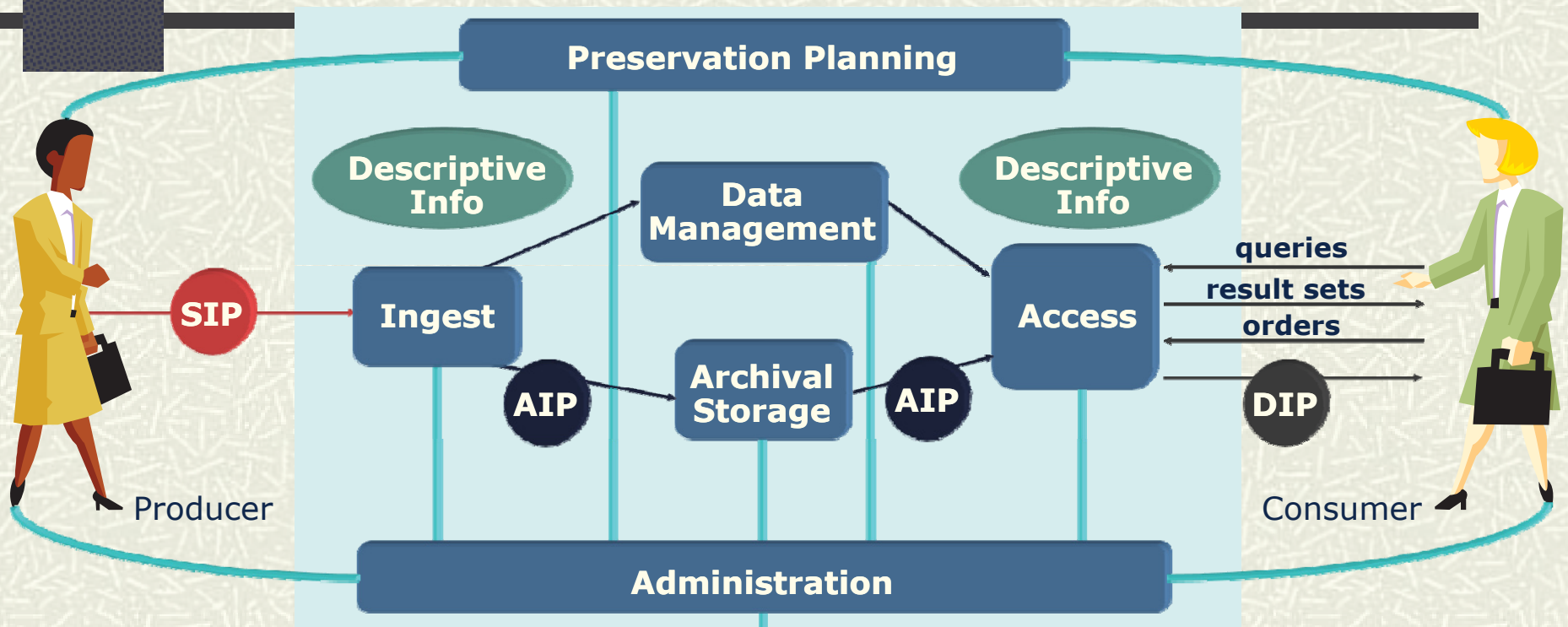
Building Preservation Repositories – Structural and Collaborative Models

- # OAIS
 - # InterPARES
 - # Cedars
 - # NEDLIB
 - # Pandora
 - # California Digital Library
-

Open Archival Information System (OAIS) Reference Model

- # Developed by the space data community
 - # High-level reference model delineating the processes and terminology to be used in the archiving of information objects
-

OAIS Functional Entities



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



Management

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



InterPARES: Strategies for Preserving Permanent Authentic Electronic Records

- # there must be a trusted archival custodian able to attest to the authenticity of copies of inactive or preserved electronic records by:
 - maintaining controls over records transfer, maintenance, and reproduction;
 - retaining documentation of reproduction processes; and,
 - capturing, through archival descriptive metadata, any changes the records have undergone since they were first created.

 - # specific preservation methods must be identified for different kinds of digital components of electronic records, e.g., a stylesheet associated with an e-mail message
-

Strategies for Preserving Permanent Authentic Electronic Records [2]

- # information must be maintained about the original form of the records and the methods needed to translate between the stored digital components and the copy of the record presented for use;

 - # preservation controls (i.e., actions, conditions, and constraints) must be established that ensure preservation and continued authenticity. There are two different kinds of preservation controls:
 - systematic controls ensure that records remain unchanged when stored offline, or during hardware or software upgrades; and,
 - dynamic controls ensure that records remain authentic when sent between people, systems, or applications.
-

Strategies for Preserving Permanent Authentic Electronic Records [3]

- # Thoroughly document the entire process of preservation, including information about the:
 - record creator's practices to support a presumption of authenticity;
 - process of bringing the records into the archives and maintaining them over time;reproduction of records.

InterPARES 1 Preservation Domain Outcomes

Conceptual outcomes:

- It is not possible to preserve an electronic record. It is only possible to preserve the ability to reproduce an authentic copy of an electronic record
- The level of functionality and other aspects of the record that need to be preserved and reproduced are decided based upon application of the authenticity requirements. The preservation method is then selected according to a risk and economic assessment
- Notion of record components

Products:

- Metamodels delineating and decomposing the activities involved in the preservation process that can be situated within the OAIS framework and can accommodate a range of preservation methods and production of different kinds of copies
-

Preservation and Metadata: Big Questions

- # Creating and keeping appropriate metadata is critical to the preservation of digital objects and the demonstration of their continued trustworthiness
 - # How can the links between information objects and metadata retain their referential integrity over time in the face of systems obsolescence, data migration, and evolution of metadata schema?
 - # Can the processes associated with ingesting and archival management of information objects generate metadata automatically?
 - # How can the trustworthiness of preserved and preservation metadata be determined in terms of quality and completeness in and over time?
-

And a Few More...

- # What metadata need to be created, when, and by whom?
 - # What metadata need to be preserved, when, by whom, and how? Conversely, what can we afford to lose over time?
 - # How much metadata is enough?
-

Cedars (curl exemplars in digital archives)

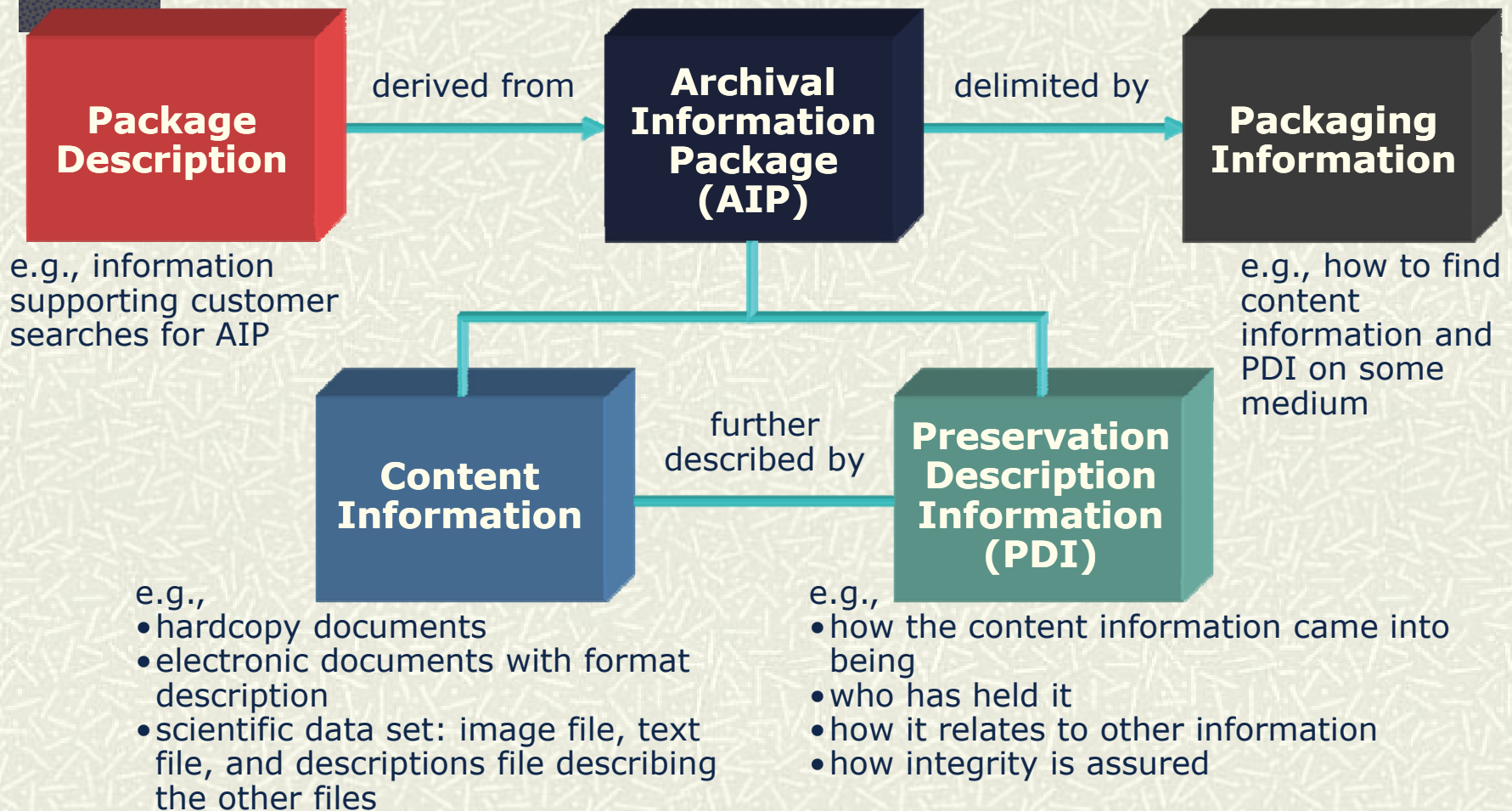
- # Primary purpose is to provide the Cedars Demonstrator Project with a basic set of preservation metadata elements to implement as part of a pilot digital archive which will be based on the Demonstrator Project system and architecture
 - # Populates the OAIS structure with metadata elements
 - # Divides an Information Package into Content and Preservation Description Information
 - # Preservation Description Information is subdivided into Reference, Context, Provenance, and Fixity information
 - # Content Information is subdivided into the Data Object and its Representation Information
 - # Each area is further subdivided to cover aspects such as Rights Management, Original Technical Environment, and Custody History
-

Information Package Definition

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



OAIS Archival Information Package



OCLC/RLG Metadata Framework to Support the Preservation of Digital Objects

- # A synthesis of preservation metadata schemes developed by OCLC, National Library of Australia, Cedars, and NEDLIB
 - # Based on OAIS Reference Model
-