open.michigan

UNIVERSITY OF MICHIGAN

SI 615 Digital Libraries Seminar

Week 6 – Interface and Infrastructure

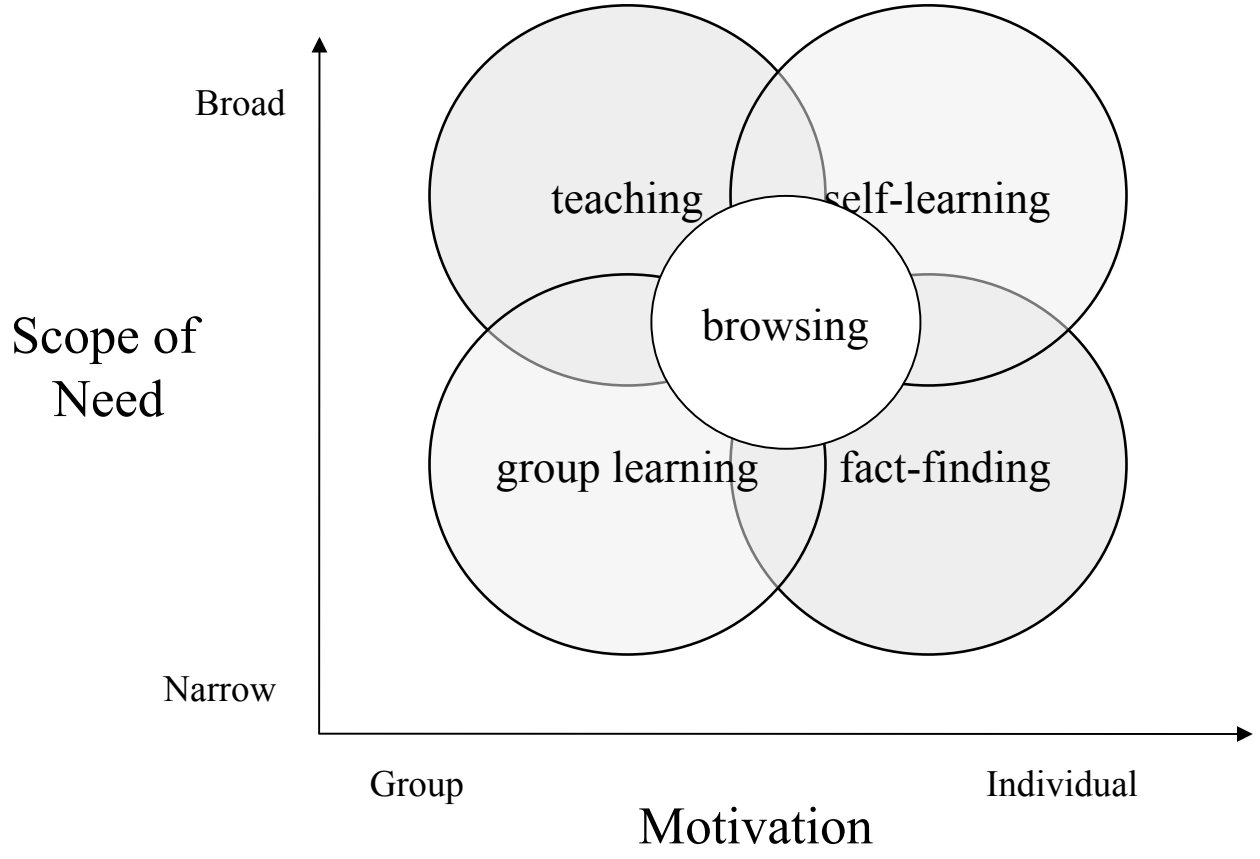
Outline of Issues

- Borgman on more useful digital libraries
- A user orientation
- Infrastructure elements
- OAIS
- OAI
- Shibboleth

Borgman's Agenda

- Metadata to data
- Independent to linked systems
- Searching to navigation
- Individual to group processes

User Domains



Academic Platform

- **Functions (Library View)**
  - Discovery: Scholars Portal discovery tool
  - Capture: harvesting and delivery tools
  - Manipulation: text-processing and citation-management tools
  - Distribution: contribution and publication tools
  - Consultation: access to virtual reference services and electronic scholarly communities

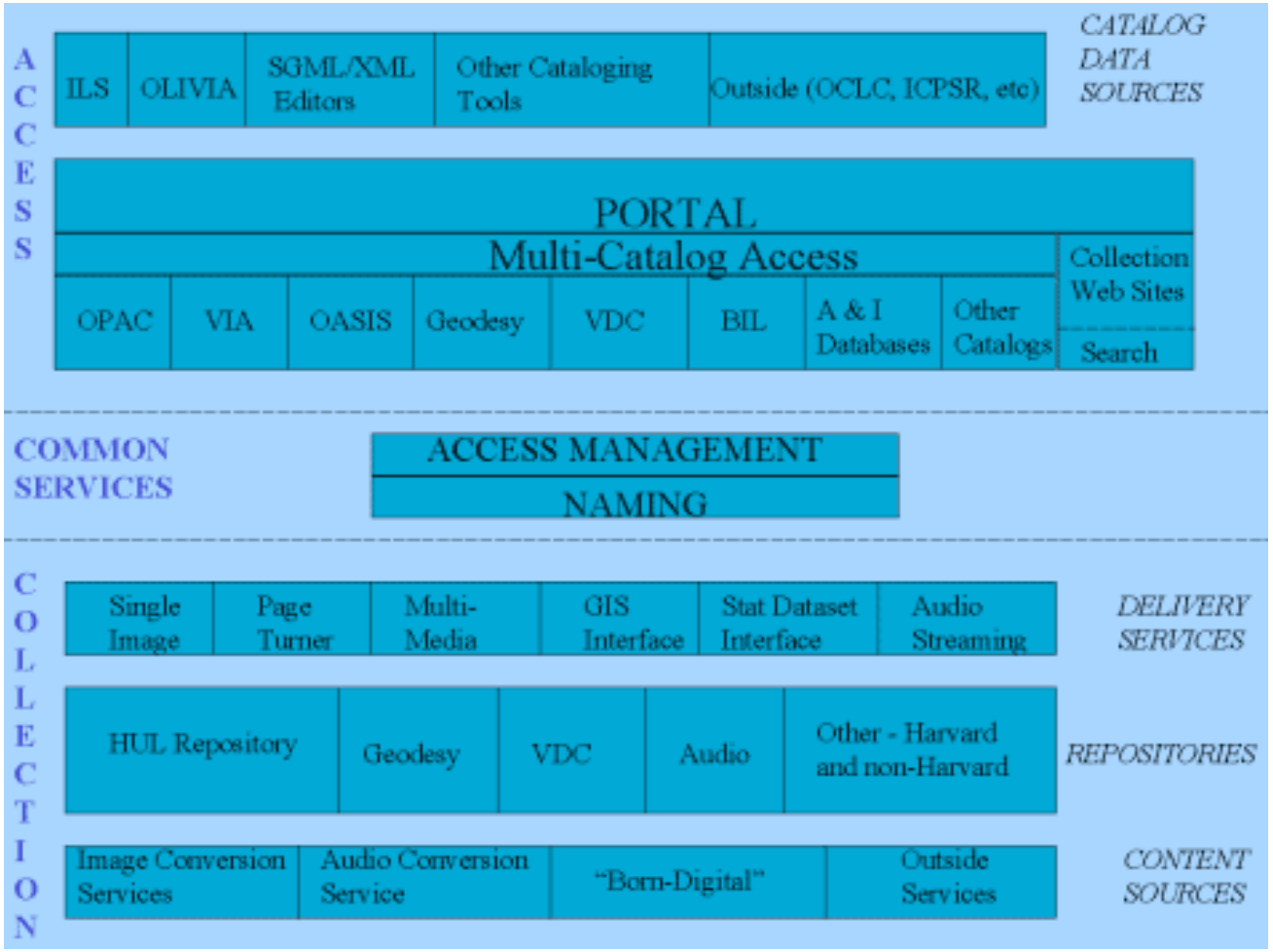
What Do Users Do?

- Discover relevant information anyplace
- Capture to their personal workspace
- Consult experts or engage scholarly communities

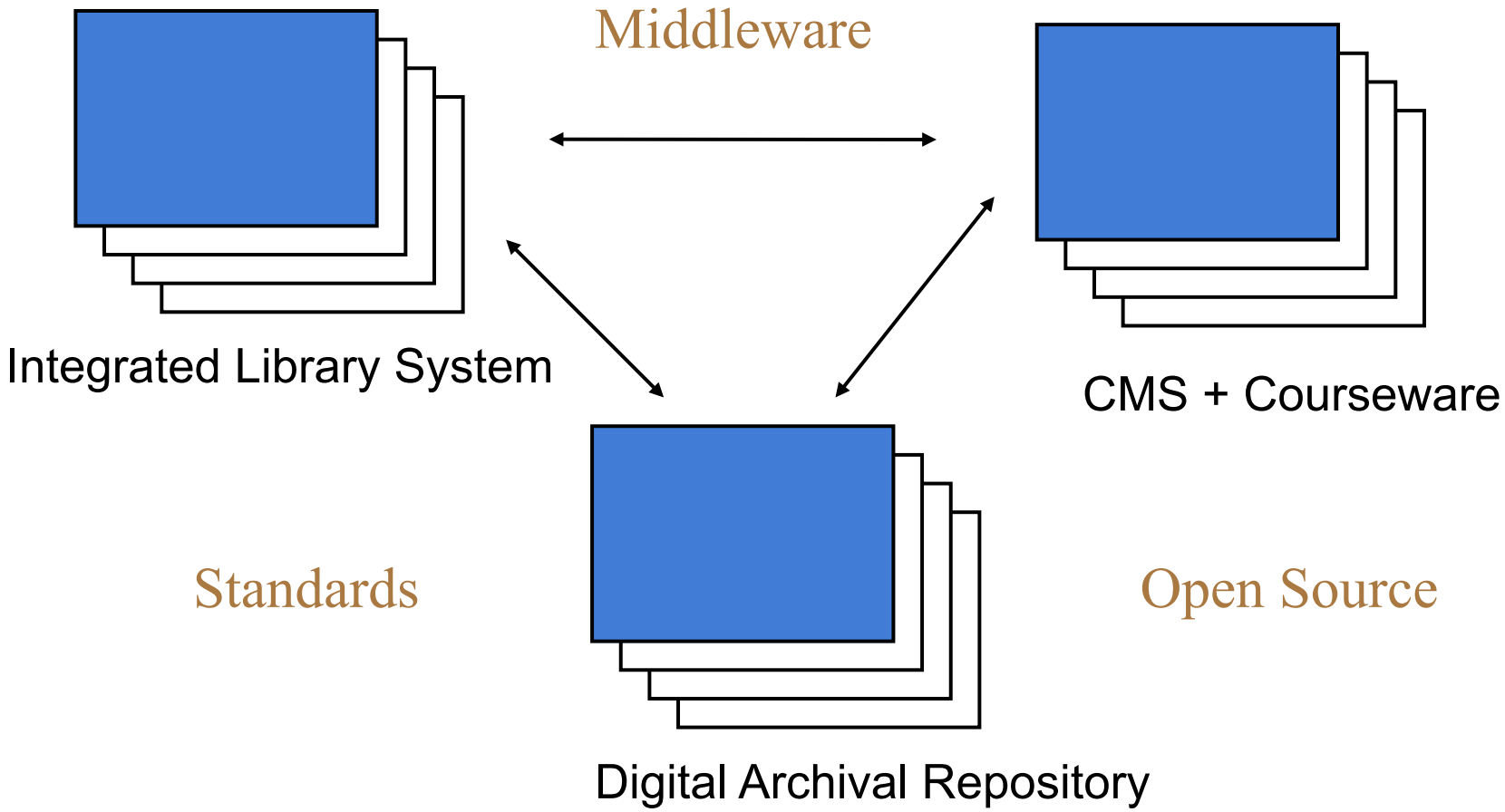
- Organize content (by concept, time, space, format)
- Cite works to track origins

- Manipulate found content (text and image)
- Distribute or publish to Web/paper

Harvard Library Digital Initiative Model



Systems Integration

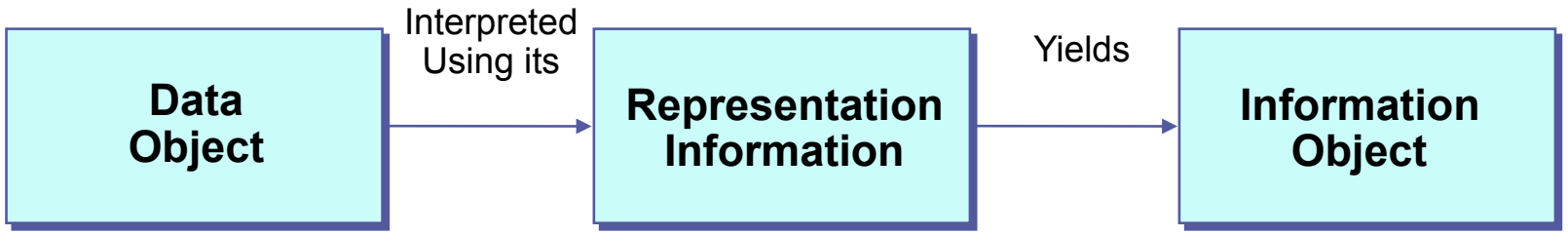


Open Archival Information System

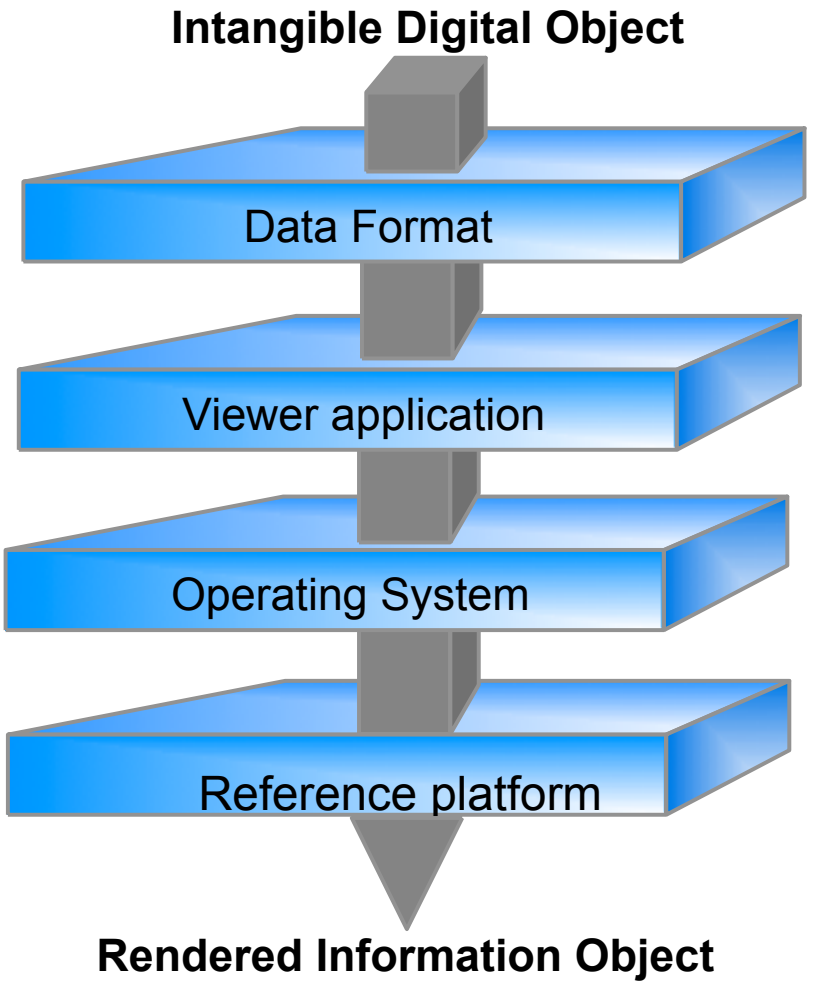
- Open
  - Reference Model standard(s) are developed using a public process and are freely available
- Information
  - Any type of knowledge that can be exchanged
  - Independent of the forms (i.e., physical or digital) used to represent the information
  - Data are the representation forms of information
- 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

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**

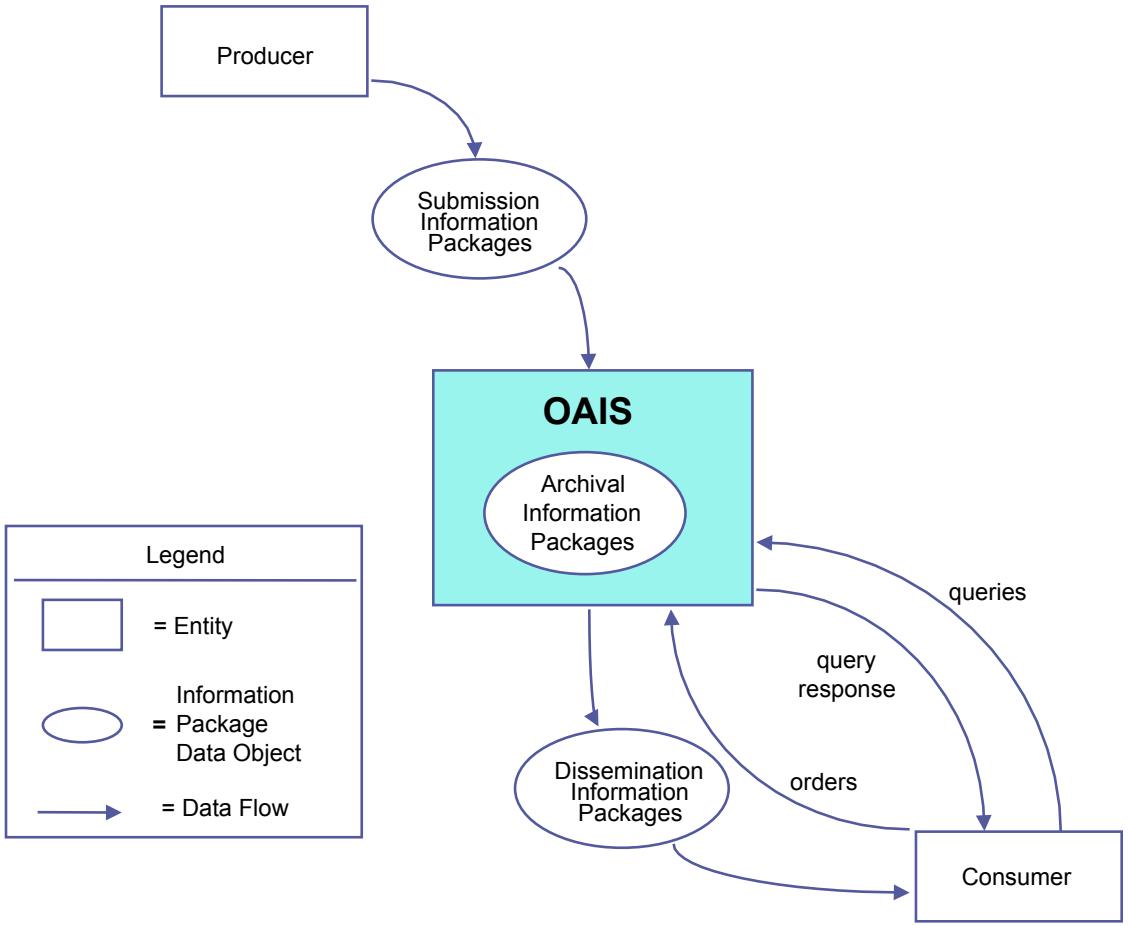


On a abstract level a PLM roughly identifies 4 abstraction levels

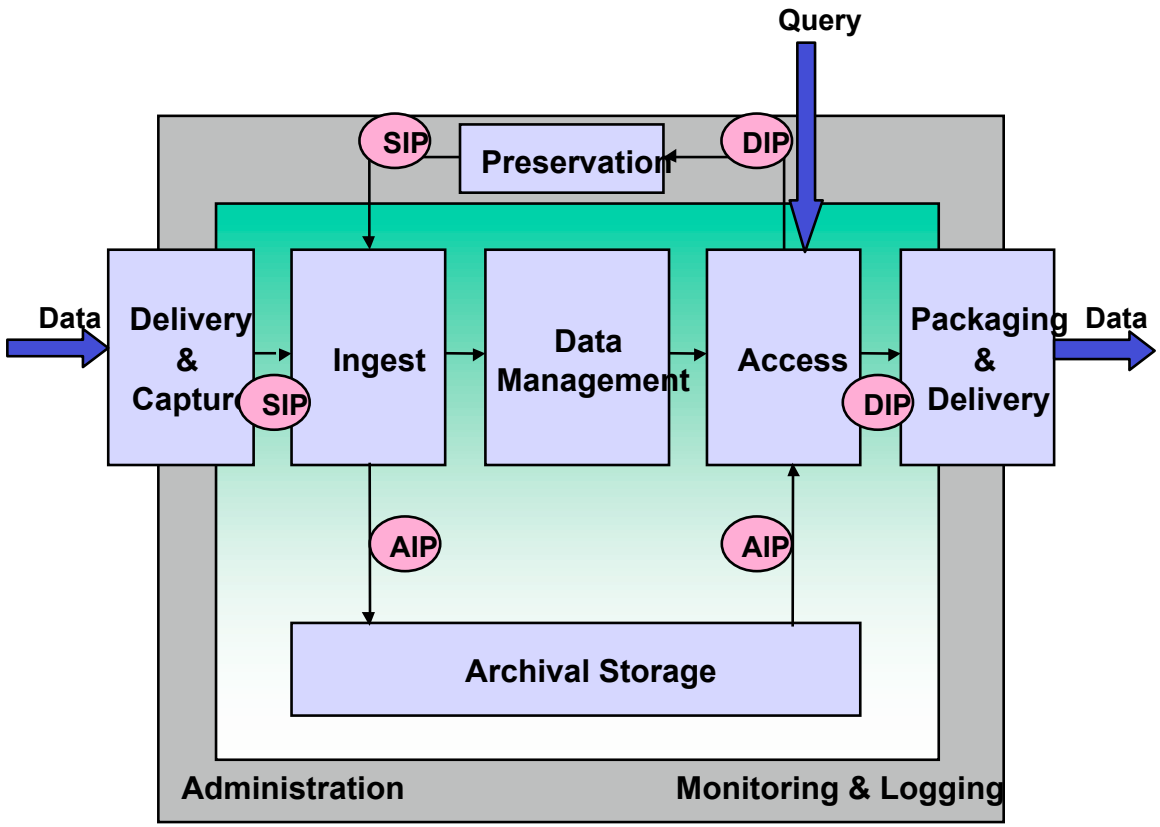


- **Data format** identifies the structuring and meaning of raw bit stream, i.e. the intangible digital object.
- The structuring and meaning of the the raw bit stream are defined within the application logic of specific **viewer applications**. These applications are used to create, modify, and present the information in its intended format.
- The **operating system** provides the shared functionality needed by all viewer applications like peripheral access and basic file management
- The **reference platform** represents the hardware on which the intangible digital objects are rendered into real world physical objects, like for instance a print out or the screen representation.

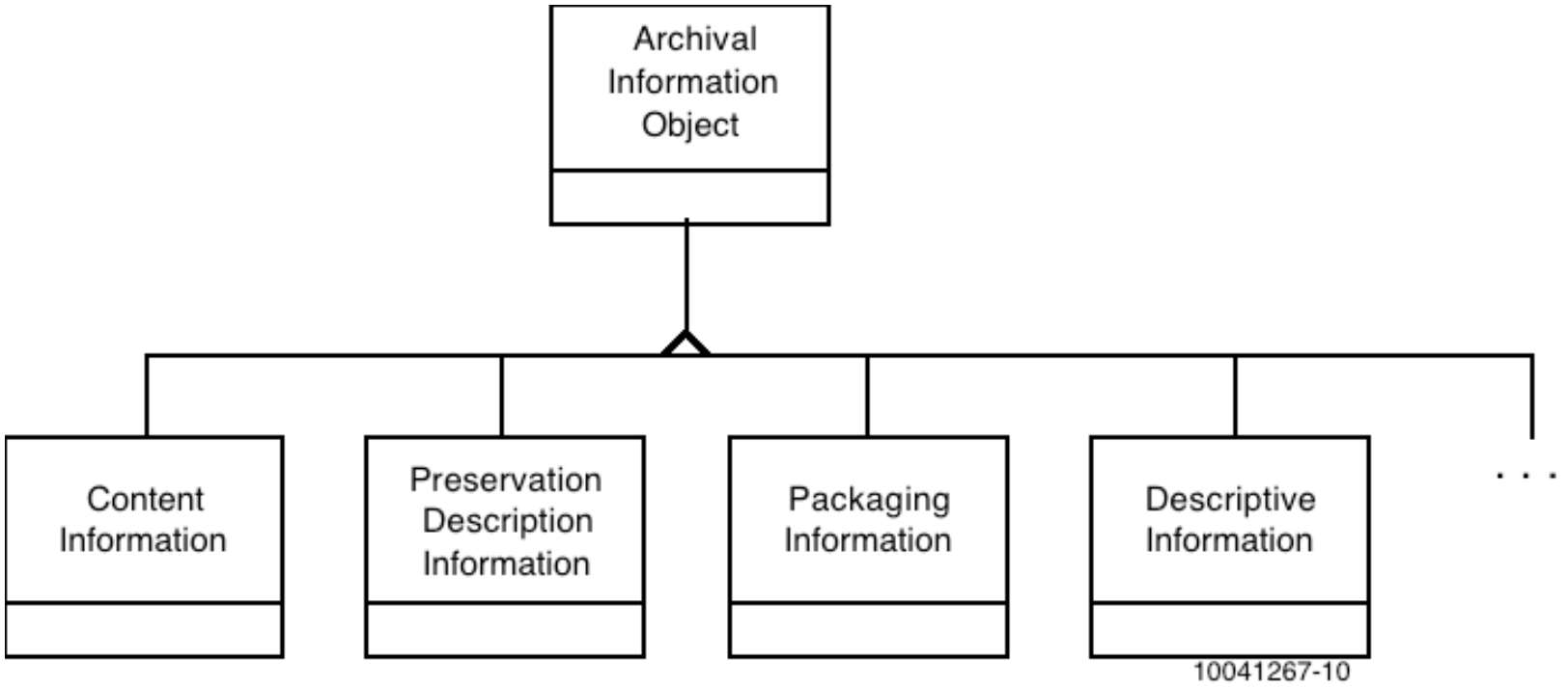
External Data Flow Diagram



Base Processes Within the OAIS Model



Types of Information Used in OAIS

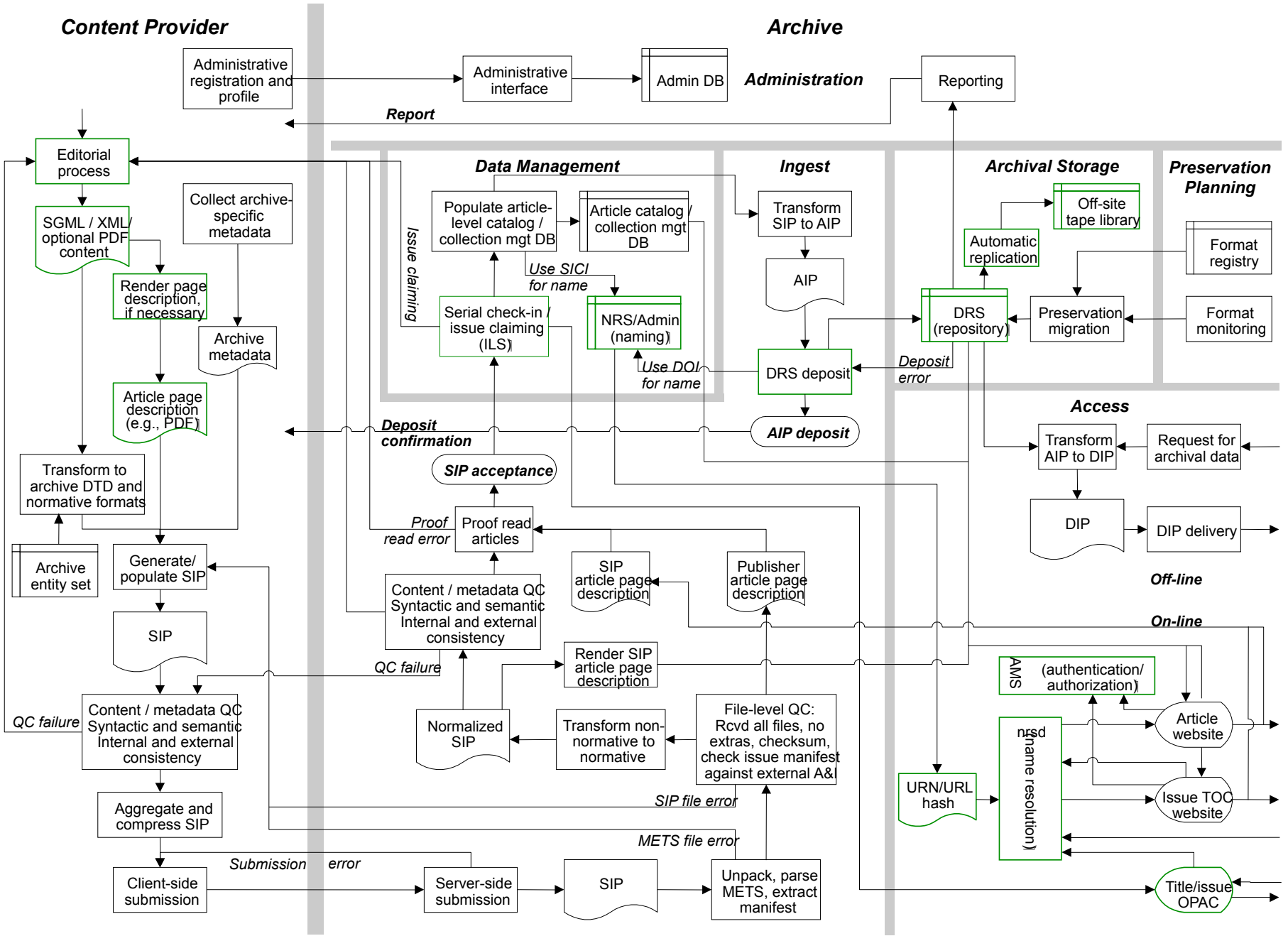


Preservation Description Information

- 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

Example of Preservation Description Information

Content Information Type	Reference	Provenance	Context	Fixity
Space Science Data	■ Object Identifier ■ Journal Reference ■ Mission, instrument, and title attribute set	■ Instrument Description ■ Processing History ■ Sensor Description Instrument ■ Instrument mode ■ Processing history ■ Decommunication map ■ Software Interface Specifications	■ Calibration history ■ Related data sets ■ Mission ■ Funding history	■ CRC ■ Checksum ■ Reed-Solomon coding
Bibliographic Information	■ ISBN ■ Title ■ Author	■ Printing history ■ Copyright ■ Position in series ■ Manuscripts ■ References	■ Related References ■ Dewy Decimal System ■ Publishing Data ■ Publisher	■ Author Digital signature ■ Cover
Software Package	■ Name ■ Author ■ Version number ■ Serial Number	■ Revision Histroy ■ License holder ■ Registration ■ Copyright	■ Help file ■ User Guide ■ Related Software ■ Language	■ Certificate ■ Checksum ■ Encryption ■ CRC

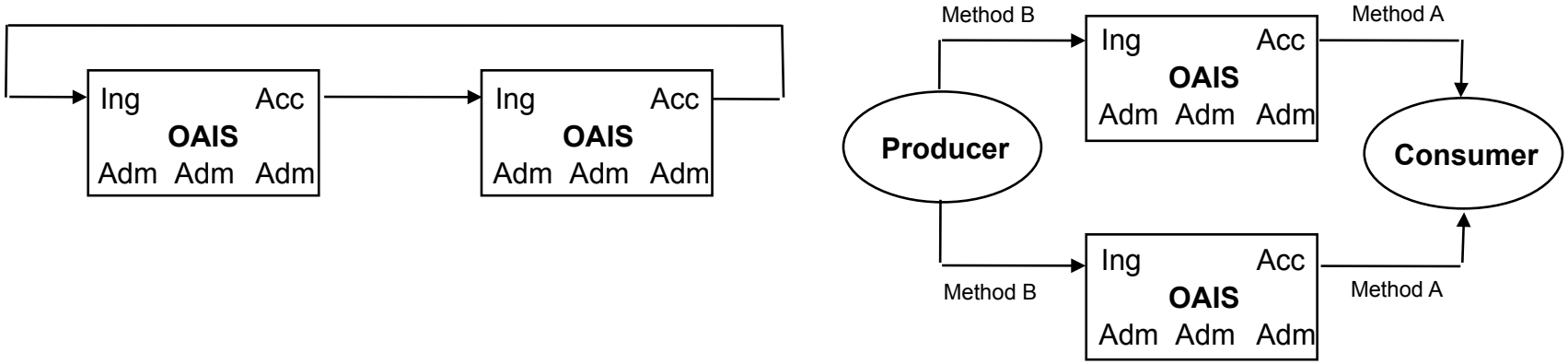


Proposed OAIS-based E-Journal Archive Mapped to Harvard's LDI Infrastructure

Categories of Archive Interactions

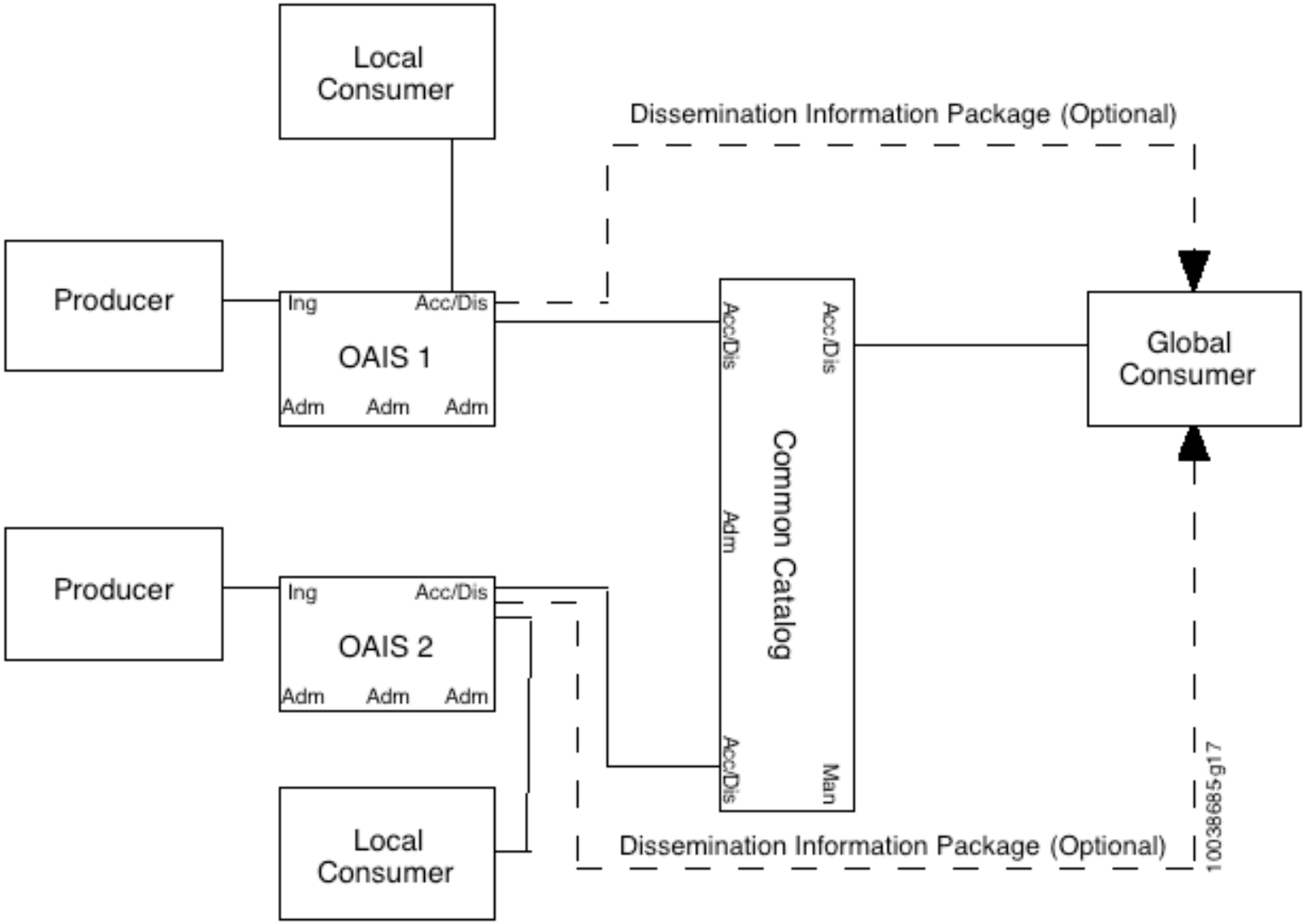
- **Independent: no knowledge by one OAIS of Standards implemented at another**
- **Cooperating: Potentially common submission standards, and common dissemination standards, but no common access. One archive may make subscription requests for key data at the cooperating archive**
- **Federated: Access to all federated OAIS is provided through a common set of access aids that provide visibility into all participating OAISs. Global dissemination and Ingest are options**
- **Shared resources: An OAIS in which Management has entered into agreements with other OAISs is to share resources to reduce cost. This requires various standards internal to the archive (such as ingest-storage and access-storage interface standards), but does not alter the community's view of the archive**

Cooperating Archives

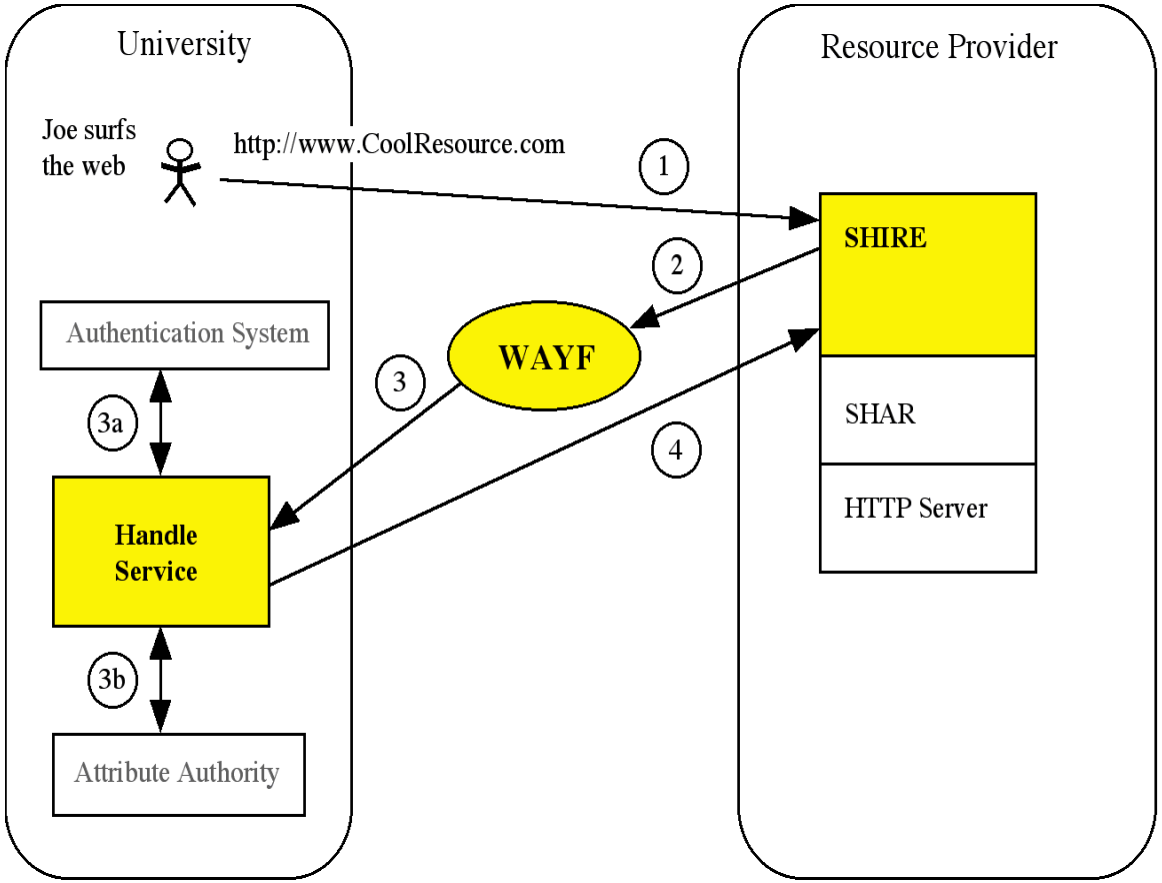


- **The first set of cooperating OAIS merely have an agreement to share at least on common SIP and DIP format to enable the transfer of holdings**
- **The second set of cooperating OAIS have standardized their DIP and SIP formats for use by producers and consumers**

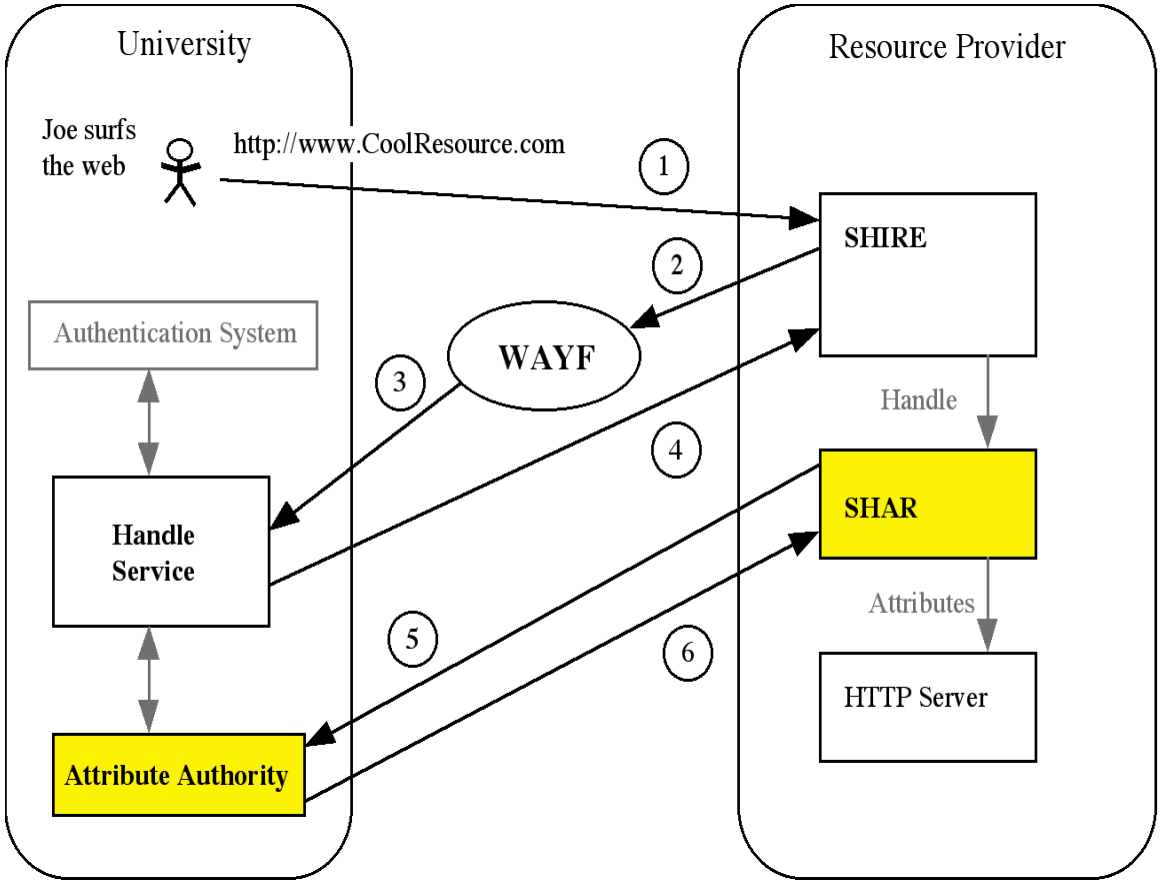
Federated Archives



Establishing a User Context



Getting Attributes and Determining Access



Attribute Authority -- Management of Attribute Release Policies

The AA provides ARP management tools/interfaces.

Different ARPs for different targets

Each ARP Specifies which attributes and which values to release

Institutional ARPs (default)

- *administrative default policies and default attributes*
- *Site can force include and exclude*

User ARPs managed via "MyAA" web interface

Release set determine d by "combining" Default and User ARP for the specified resource

Authorization Attributes

Typical Attributes in the Higher Ed Community

Affiliation	"active member of the community"	Member@washington.edu
EPPN	identity	gettes@georgetown.edu
Entitlement	An agreed upon opaque string	Urn:mace:infovendor:contract1234
OrganizationalUnit	Department	Economics Department
EnrolledCourse	Opaque course identifier	Physics 201

Thank you!

Paul Conway
Associate Professor
School of Information
University of Michigan
www.si.umich.edu