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

- Scope of Need
  - Narrow
  - Broad

- Motivation
  - Group
  - Individual

- Fact-finding
- Teaching
- Self-learning
- Browsing
- Group learning
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

Integrated Library System

Middleware

CMS + Courseware

Standards

Open Source

Digital Archival Repository
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
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

Legend

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>□</td>
<td>Entity</td>
</tr>
<tr>
<td>○</td>
<td>Information Package</td>
</tr>
<tr>
<td>▽</td>
<td>Data Object</td>
</tr>
<tr>
<td>→</td>
<td>Data Flow</td>
</tr>
</tbody>
</table>

Producer

Submission Information Packages

OAIS

Archival Information Packages

Dissemination Information Packages

queries

query response

orders

Consumer
Base Processes Within the OAIS Model

- Ingest
- Data Management
- Archival Storage
- Delivery & Capture
- Packaging & Delivery
- Administration
- Monitoring & Logging

- Data flow:
  - Data input to Delivery & Capture
  - Ingest
  - Data Management
  - Access
  - Packaging & Delivery

- Data types:
  - SIP
  - DIP
  - AIP

- Query:
  - From outside to Preservation
  - From Access to Packaging & Delivery

- Archival Storage:
  - Underlying storage for data management and access.
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

<table>
<thead>
<tr>
<th>Content Information Type</th>
<th>Reference</th>
<th>Provenance</th>
<th>Context</th>
<th>Fixity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Science Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Object Identifier</td>
<td>▪ Instrument Description</td>
<td>▪ Calibration history</td>
<td>▪ CRC</td>
</tr>
<tr>
<td></td>
<td>▪ Journal Reference</td>
<td>▪ Processing History</td>
<td>▪ Related data sets</td>
<td>▪ Checksum</td>
</tr>
<tr>
<td></td>
<td>▪ Mission, instrument, and title attribute set</td>
<td>▪ Sensor Description</td>
<td>▪ Mission</td>
<td>▪ Reed-Solomon coding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Instrument</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Instrument mode</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Processing History</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Decommunication map</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Software Interface Specifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bibliographic Information</td>
<td>ISBN</td>
<td>▪ Printing history</td>
<td>▪ Related References</td>
<td>▪ Author Digital signature</td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>▪ Copyright</td>
<td>▪ Dewy Decimal System</td>
<td>▪ Cover</td>
</tr>
<tr>
<td></td>
<td>Author</td>
<td>▪ Position in series</td>
<td>▪ Publishing Data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ Manuscripts</td>
<td>▪ Publisher</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>▪ References</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software Package</td>
<td>Name</td>
<td>▪ Revision History</td>
<td>▪ Help file</td>
<td>▪ Certificate</td>
</tr>
<tr>
<td></td>
<td>Author</td>
<td>▪ License holder</td>
<td>▪ User Guide</td>
<td>▪ Checksum</td>
</tr>
<tr>
<td></td>
<td>Version number</td>
<td>▪ Registration</td>
<td>▪ Related Software</td>
<td>▪ Encryption</td>
</tr>
<tr>
<td></td>
<td>Serial Number</td>
<td>▪ Copyright</td>
<td>▪ Language</td>
<td>▪ CRC</td>
</tr>
</tbody>
</table>
Provisional OAIS-based E-Journal Archive Mapped to Harvard’s LDI Infrastructure

Content Provider

- Administrative registration and profile
- Editorial process
  - SGML / XML optional PDF content
  - Render page description (e.g., PDF)
- Collect archive-specific metadata
- Render article page description (e.g., PDF)
- Archive metadata
- Transform to archive DTD and normative formats
- Generate/populate SIP
- Issue claiming
  - Request for archival data
    - Off-site tape library
    - Request for archival data
  - On-line access
    - Article website
      - Title/issue OPAC
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
    - NRS/Admin (naming)
      - Automatic replication
    - DIP
      - DIP delivery
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
Establishing a User Context


1. SHIRE
2. SHAR
3. Handle Service
4. Attribute Authority

Joe surfs the web

Authentication System

WayF
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 determined by “combining” Default and User ARP for the specified resource
Authorization Attributes

Typical Attributes in the Higher Ed Community

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliation</td>
<td><a href="mailto:Member@washington.edu">Member@washington.edu</a></td>
</tr>
<tr>
<td>EPPN</td>
<td><a href="mailto:gettes@georgetown.edu">gettes@georgetown.edu</a></td>
</tr>
<tr>
<td>Entitlement</td>
<td>Urn:mace:infovendor:contract1234</td>
</tr>
<tr>
<td>OrganizationalUnit</td>
<td>Economics Department</td>
</tr>
<tr>
<td>EnrolledCourse</td>
<td>Physics 201</td>
</tr>
<tr>
<td>“active member of the community”</td>
<td><a href="mailto:Member@washington.edu">Member@washington.edu</a></td>
</tr>
<tr>
<td>identity</td>
<td><a href="mailto:gettes@georgetown.edu">gettes@georgetown.edu</a></td>
</tr>
<tr>
<td>An agreed upon opaque string</td>
<td>Urn:mace:infovendor:contract1234</td>
</tr>
<tr>
<td>Department</td>
<td>Economics Department</td>
</tr>
<tr>
<td>Opaque course identifier</td>
<td>Physics 201</td>
</tr>
</tbody>
</table>
Thank you!

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