Digital Library Federation

Forum on Digital Library Practices
Washington, DC, July 17-18, 1999

Briefing Session 3 — Digital Repositories

Jim Coleman
Harvard University Library
jim_coleman@harvard.edu

Why Do I Need A Repository?

- Digital Objects Are More Numerous/ Volatile/Mutable Than Predecessors
- Digital Objects Depend On and Are Bound To Technical Environment/Infrastructure
- Digital Libraries Share Similar Levels of User Service Needs with “Traditional” Libraries
- No More “404” Errors

What *IS* A Repository, Anyway?

- A Set of Services
  - Naming Services
  - Object Management
  - Object Discover/Recovery
  - Object Security
  - Object Migration
  - Reporting
- A Physical/Virtual Datastore

Working Assumptions

- Repository services will manage the user's bits
- Repository services will return those bits to the owner or applications trusted by the owner (aka "trusted applications" / disseminators)
- No assumptions about what the digital object **is**
- Repository stands in support of Harvard's LDI Project

A Word or Two About LDI

- Create Infrastructure for the Library's Electronic Components
- Create "advice" services for the creators and disseminators of digital resources
- Create applications (catalogs and other types of applications) to manage digital objects
- Fund an internal grant program in support of the creation and delivery of digital objects

Object Management

- Who Owns the Object
- What Does the Object Consist Of
- How Does the Object Relate to Other Objects
- How Is the Object Created/Updated
- What Applications Have Access to the Object

Relational Model for Object Management

BASIC METATA

ID
URN
LOCAL_NAME
OWNER
APPS
BYTES
MIME_TYPE
STATUS
LOCATION

OBJECT

ID
BLOB

INVENTORY

ID
MD_REC
TYPE

OTHER TYPES

IMAGE
MOAII
DC

TRUSTED APPLICATIONS

APPLICATION
CONTEXT (METHOD)

Objects & Applications Are Opaque/Services Are Not

- Repository Does Not “Model” Complex Objects, Applications Do
- Repository Services Assume a Simple Fetch Model
- Complex Interactions with Other Services (Naming, Authentication, Accounting)
- Applications Answer the “What is an Object?” Question

Current “Trusted Applications”

- Naming and Repository Services Themselves
- Digital Table of Contents Project
- Ereserves
- VIA
- Other LDI Funded Projects

Object Security

- Who Owns the Object
- Who Is Entitled to the Object
  - Which Users
  - Which Applications
  - Which Components
- Is the Object “Authentic”?
- Is **MY** Object Being Misused by Others?

Reporting

- Reports on:
  - object use
  - object size/total storage
  - success/failure of input/update and batch operations
  - billing

User Services/Management

- User Account Creation
- FTP sites for batch data/metadata transfer
- Varying Batch Input Models (XML, CSV)
- Batch Transaction Processing/Tracking
- Other Services Built Out of Current "Trusted Applications"

Current Implementation Environment

- Oracle 8.1.5 as DBMS
- Modified Apache server for Naming services
- Load utilities
  - Web Forms
  - Batch (via SQLLoader, PL/SQL) w/ pre-processing routines in Java/perl
- Environment pushes Oracle services/envelope

Futures

- Migration
- Object Discovery/Recovery
- Relational to OO Data Model

Object Migration

- TIFF Today/PNG Tomorrow
- “Archival Quality” Digital Object vs.. “Latest/Greatest” Delivery Surrogate
- Object “Refreshment”
- Object “Aging” — When Do I Need to Redigitize … and Why

Object Discover/Recovery

- Relationship Between Object Components and “Whole Digital Object”
- Relationships That Encourage Object Reuse in Novel Ways
- Relationship Between Object Metadata / Hollis Catalog Record
- Object “Reveals” Itself

Relational to OO Data Model

- Model services with data object
  - beyond simple “fetches”
  - “intelligent” interfaces to object
- Model object with object’s metadata
- Small set of clearly defined objects, based on community standards
- Relationship of **catalogs** to **applications**