The Digital Library Federation Aquifer Initiative

Jon Dunn and Jenn Riley
Indiana University Digital Library Program

IU Digital Library Brown Bag Series
February 27, 2008
Outline

- Background and motivations for DLF Aquifer
- American Social History Online Portal
- Metadata and OAI harvesting
- Portal technology and implementation scenarios
- Conclusion
Background

- Digital Library Federation
  - Consortium of libraries and allied institutions
  - Mission: enable new research and scholarship of its members, students, scholars, lifelong learners, and the general public by developing an international network of digital libraries

- DLF Aquifer
  - Based on goal in DLF charter to implement a “distributed, open digital library” built from collections across DLF member (and other) libraries and archives
  - Began in 2003; really got going in 2005
DLF Aquifer

- **Purpose**
  - “To promote effective use of distributed digital library content for teaching, learning, and research in the area of American culture and life” by making digital content available to scholars and students where they do their work

- **Activities**
  - Developing systems, schemas and protocols
  - Developing and promoting best practices
IU Involvement in DLF Aquifer

- Working groups:
  - Metadata
    - Jenn Riley (member and current chair)
  - Technology/Architecture
    - Jon Dunn (member and past chair)
  - Services
    - John Walsh (former member)
- Currently a subcontractor on American Social History Online Mellon grant
American Social History Online (ASHO)

- Funded by grant to DLF Aquifer from the Andrew W. Mellon Foundation: *DLF Aquifer Development for Interoperability Across Scholarly Repositories*
  - April 2007 – March 2009
- Develop a system for scholars to use **distributed** digital collections related to 19\(^{th}\) and 20\(^{th}\) century American social history through **consistent** interfaces
- System demonstrates technologies and architecture that can be reused in other domains
Demonstration of ASHO Portal

- www.dlfaquifer.org
Architecture Overview

Digital library repositories

Selection and metadata remediation

Aggregated Metadata Store

OAI-PMH & MODS

Web Portal

SRU Target

OAI-PMH Data Provider

Other Protocols/APIs

Local tools and services
Architecture Overview

Digital library repositories

Selection and metadata remediation

Aggregated Metadata Store

OAI-PMH & MODS

Web Portal

SRU Target

OAI-PMH Data Provider

Other Protocols/APIs

Local tools and services
Metadata for ASHO portal

- Much work has been done on metadata aggregations in the recent past
- Biggest lesson is that they’re harder than we thought!
- Aquifer provided a forum for bringing together and documenting best practices in this area
- Work began before Mellon project funded, but principles should apply to many types of metadata aggregations
Bringing together metadata for distributed collections

- Aquifer chose to initially use OAI-PMH to collect metadata for the test portal - this is not necessarily the only or a permanent solution
- OAI-PMH allows metadata in any format expressed in an XML Schema to be shared
- Aquifer Metadata Working Group recommended metadata for the project be harvested in the MODS format
  - Working group members had first-hand experience with the limitations of simple Dublin Core in metadata aggregations
  - Expectation that DLF members are among those most capable of exposing MODS for harvesting
Challenges for using MODS

- MODS is an inherently hierarchical format, making it difficult for systems based on relational databases to support it well.
- Most OAI-PMH data provider software still does not natively support exposing MODS.
- Few metadata harvesters collect MODS so there has been little need for data providers to do the work to provide it.
- Lack of experience with using MODS for metadata in aggregations.
Aquifer MODS Guidelines

- Metadata for aggregation is different than metadata for local discovery and management, so metadata guidelines were needed
- Metadata Working Group members brought to the project experience as both aggregators and data providers
Structure of MODS Guidelines

- One entry per top-level MODS element
  - Summary of Requirements
  - Definition from MODS User Guidelines
  - Discussion of Use
  - Examples
  - Use by Aggregators
  - Mapping to Dublin Core
  - Relationship to DLF/NSDL Best Practices for Shareable Metadata
MODS Guidelines Levels of Adoption

- MODS Guidelines are a big scary document - we needed a more accessible introduction to the metadata requirements
- Presents a user-centric, rather than record-centric, view of metadata requirements
- Defined five levels based on functions meeting that level supports
The five levels

- **Minimum for participation:** Allows users to cite the resource
- **Minimum for doing anything useful:** Allows users to perform basic searches and filtering
- **Allows more advanced functionality:** Allows users to browse and group search results
- **Adopt all required guidelines (and some recommended):** Allows users to perform more precise searches
- **Completely adopt all recommendations:** Allows users to effectively evaluate resources
MODS Guidelines FAQ

- Quick way for the Metadata Working Group to release additional information without revising the Guidelines
- Document answers to commonly-asked questions
- Respond to changes in MODS without revising the Guidelines
MODS evaluation service

- Hope to have available soon
- Metadata working group translating Guidelines into Schematron rules organized by Level of Adoption
- Will be used as the basis for a Web form that accepts an uploaded MODS file and provides an HTML report describing conformance to the guidelines
Architecture Overview

Digital library repositories

OAI-PMH & MODS

Selection and metadata remediation

Aggregated Metadata Store

Web Portal

SRU Target

OAI-PMH Data Provider

Other Protocols/APIs

Local tools and services
Portal Context

- Web portal is not the only delivery mechanism for Aquifer
- Need to make content available to users where they are already doing their work – “in the flow”
- System architecture provides ways to reexpose aggregated metadata and collections to other tools and services
  - Searching of metadata via SRU and OpenSearch
  - Reharvesting of item-level and collection-level metadata via OAI-PMH
Local Implementation Scenarios

- Commercial search service
  - Examples: Google, Yahoo
- Citation management tools
  - Examples: Zotero, RefWorks, Endnote
- Federated search tool
  - Examples: SirsiDynix SingleSearch, ExLibris MetaLib
- Course management system
  - Examples: Sakai, Blackboard, Angel
IU Local Implementation

Scenario: Sakai/Oncourse

- Integrate access to Aquifer content through Oncourse for faculty and students
- Build on the Citations Helper tool developed by the Sakaibrary project
- Two phase approach
  - 1. Basic searching
  - 2. Content reuse
- Demonstration
Asset Actions

- Sharing of descriptive metadata (e.g. via OAI-PMH) is widespread
- How to expose deeper capabilities of digital objects?
  - Examples: versions of an image, pages of a book, scenes of a video
- Similar to Fedora notion of “behaviors”
- Hypothesis:
  - Not everyone can/will implement Fedora
  - Existing approaches (e.g. METS profiles) too difficult to agree on
- Goal: simple, low-barrier-to-entry method of exposing digital library content and structure
Asset Action Groups

- Define sets of actions corresponding to various types of digital objects
  - Default
  - Basic Image
  - Addressable Image
  - Structured Text
  - Image Text (Paged Image)
Asset Action Groups:

Examples

**Default**
- getAssetDefinition
- getPreview
- getLabel
- getDCRecord
- getWebView
- getDefaultContent

**Basic Image**
- getThumbnail
- getScreenSize
- getMaxSize
- getDynamicView
<aqa:assetActions type="image" objid="2559647" xmlns:aqa="http://dca.tufts.edu/ad"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://dca.tufts.edu/ad
   http://gita.grainger.uiuc.edu/AquiferTechWG/assetactions-v1.3b.xsd">
  <aqa:actionGroup label="default Action Group" name="defaultActionGroup">
    <aqa:action label="Asset Definition" name="getAssetDefinition" format="text/xml"
               url="http://sulzer.at.northwestern.edu/tomcat/cocoon/oai/posters_wii Assetdef/2559647" />
    <aqa:action label="Preview" name="getPreview" format="image/gif"
               url="http://www.library.northwestern.edu/govpub/collections/wii-posters/img/ww0207-77-sml.gif" />
    <aqa:action label="Label" name="getLabel" format="text/plain"
               url="http://sulzer.at.northwestern.edu/tomcat/cocoon/oai/posters_wii-label/2559647" />
    <aqa:action label="Description" name="getDCRecord" format="text/xml"
               url="http://sulzer.at.northwestern.edu/tomcat/cocoon/oai/posters_wii-dc/2559647" />
    <aqa:action label="Web View" name="getWebView" format="text/html"
               url="http://www.library.northwestern.edu/otcgi/digilib/liscgi60.exe?DB=2&ACTION=View&QUERY=2559647&OP=and&SUBSET=SUBSET&FROM=1&SIZE=5&ITEM=1" />
    <aqa:action label="Default Content" name="getDefaultContent" format="text/html"
               url="http://www.library.northwestern.edu/otcgi/digilib/liscgi60.exe?DB=2&ACTION=View&QUERY=2559647&OP=and&SUBSET=SUBSET&FROM=1&SIZE=5&ITEM=1" />
  </aqa:actionGroup>
  <aqa:actionGroup label="basic Image Action Group" name="basicImageActionGroup">
    <aqa:action label="Thumbnail" name="getThumbnail" format="image/gif"
               url="http://www.library.northwestern.edu/govpub/collections/wii-posters/img/ww0207-77-sml.gif" />
    <aqa:action label="Screen Size" name="getScreenSize" format="image/jpeg"
               url="http://www.library.northwestern.edu/govpub/collections/wii-posters/img/ww0207-77.jpg" />
    <aqa:action label="Max Size" name="getMaxSize" format="image/jpeg"
               url="http://www.library.northwestern.edu/govpub/collections/wii-posters/img/ww0207-77.jpg" />
    <aqa:action label="Dynamic View" name="getDynamicView" format="text/html"
               url="http://www.library.northwestern.edu/otcgi/digilib/liscgi60.exe?DB=2&ACTION=View&QUERY=2559647&OP=and&SUBSET=SUBSET&FROM=1&SIZE=5&ITEM=1" />
  </aqa:actionGroup>
</aqa:assetActions>
Asset Actions: Next Steps

- Adopt in ASHO portal
- Look at expression in OAI-ORE
  - Open Archives Initiative Object Reuse and Exchange
- Use in Sakai?
Future of the Aquifer initiative

- Mellon-funded ASHO project ends in March 2009
- Formal process will begin shortly to find a home for the ASHO portal
- Aquifer as an initiative will likely wrap up around the same time as the ASHO project as DLF turns its attention elsewhere
- But several more things to do between now and then…
ASHO Assessment Activities

- Led by Services Working Group
- Several approaches
  - Survey comparing overall portal and commercial search
  - Focus groups comparing commercial search and portal
  - Zotero - Portal semester longitudinal assessment
  - Overall portal and commercial search interviews (faculty)
  - Overall portal & federated search (observation – usability)
  - Sakai (observation – usability)
- Metadata Working Group will use assessment results to plan changes to MODS Guidelines and Levels of Adoption
Metadata remediation

- $18,000 from the Delmas foundation
- Review tools for metadata enhancement
  - Test their effectiveness
  - Compare benefit against need and ease of use
- Make recommendations for tools best incorporated into a production metadata harvesting workflow
Thank you!

- Questions?
- More information:
  - ASHO Portal: http://www.dlfaquifer.org
  - Aquifer Project Information Page: http://wiki.dlib.indiana.edu/confluence/x/4F4
  - DLP Brown Bag Presentation Archive: http://www.dlib.indiana.edu/education/brownbags/