Image Collection Workflow and Tools
a technical overview

Michael Durbin
2010 Brown Bag Presentation Series
April 21, 2010
Introduction and Overview

- Steps to complete a collection for publication
  - Starting from a set of digitized images and metadata model
  - Ending with a public-facing application for users to access the materials
- Tools and Applications
Starting Point: Images and Metadata Plan

**Sources**

- New digitization effort
- Migration of existing collections
- Combination: Migration coupled with new digitization
- Other
Examples

- US Steel
  - old site / new site*
- Liberian Collections Photographs
  - cataloging access
- Digital Library of the Commons: Images
  - new site
- Hohenberger
  - old site / new site* / cataloging access
WORKFLOW AND TOOLS
Workflow & Tools: Overview

- Ingest objects – Ingest Tool
- Index objects – Fedora Index Service
  - fielded search, browse, facets, ranking, sort…
- Cataloging interface - PhotoCat
  - configure fields, controlled vocabulary interaction
- Discovery and access interface
  - configure fields, search behavior, browsing, facets, controlled vocabulary interaction
Workflow: Ingest

Physical Object | Digital Files | Repository Object

Hoh009.000.0089.tif
Hoh009.000.0089-mods.xml
Hoh009.000.0089-mix.xml

iudl:5081
DC
METADATA (METS)
MODS
MIX
THUMB
SCREEN
LARGE
SCALABLE
RELS-EXT
Tools: Ingest Tool

- History
  - Started development in 2006
  - Rewritten/Adapted several times

- Functionality
  - Transform any existing descriptive metadata
  - Generate technical metadata from image files
  - Reconcile with existing objects in repository
  - Create/Update objects in repository

- Configuration
  - Specify object type, file locations, etc.
Workflow: Indexing and Searching
Tools: Fedora Index Service and SRU Server

- **History**
  - Born out of Fedora GSearch and OCLC SRW service around 2006
  - Has been extended and improved for nearly every collection

- **Architecture**
  - Receives messages from the repository
  - Maintains indexes
  - Exposes searches using SRU
Tools: Fedora Index Service and SRU Server

- Configuration
  - Fedora Index Service
    - XSLT to transform metadata into name value pairs
    - Configuration for field analysis
  - SRU Service
    - List of properties describing fields to expose
    - Field and query analysis
Tools: PhotoCat

- History
  - Developed to meet a growing need starting around 2007

- Functionality
  - fully configurable metadata model
  - fully configurable field display
  - loads of special purpose improvements

- Configuration
  - Field display
  - Field interaction
  - Field validation
  - Field storage
  - Search behavior
Workflow: Discovery Application

The Digital Library Program is proud to present the U.S. Steel Gary Works Photograph Collection, a series of more than 2,200 photographs of the Gary Works steel mill and the corporate town of Gary, Indiana held by the Cabinet Regional Archives at Indiana University Northwest. In images of compelling diversity, historians and the general public can view all aspects of this planned industrial community: the steel mill, the city, and the citizens who lived and worked there.

The Digital Library Of The Commons Image Collection

About the Collection

This digital image collection is comprised of photographs and slides taken by researchers from the Workshop in Political Theory and Policy Analysis over the past 30 years. The images were taken primarily during field trips and conference visits, and showcase diverse types of common pool resources, such as Nepal irrigation systems, African forests, conferences of the International Association for the Study of the Commons, and research sites of the International Forestry Resources and Institutions (IFRI) Program.

Digital Library Program

Indiana University

Libraries/University Information Technology Services
Workflow: Discovery and Access Interface

- History
  - Born from code written for the Slocum collection around 2006
  - Incorporated enhancements needed for IN Harmony

- Functionality
  - Search
  - Browse
  - Facets
  - Filters
  - Date support
  - CV query enhancements
  - JQA query analysis
Workflow: Discovery and Access Interface

- Configuration
  - Search fields
  - Browse fields
  - Facet fields
  - Thesaurus interaction
  - Search results fields
  - Full record fields
- Look and Feel
Workflow: Access Interface Branding

- Colors/Fonts/Icons
  - Conform to IU visual identity guidelines
  - Conform to collection manager’s vision for the collection
- Tools
  - CSS
  - STRUTS
  - Photo Render
Tools: photorender

- Uses SRU search and the repository as the source for image data
- Has several pre-configured layout patterns
- Supports user manipulation of the viewpoint
- Written in Java using the Java3D library
Example: Hohenberger Photographs

- Set up Indexer/SRU
- Set up Photocat
- Customize search and browse application

DIGITAL LIBRARY PROGRAM
INDIANA UNIVERSITY
Libraries/University Information Technology Services

April 22, 2010
Workflow Summary

😊 Great and versatile tool set
   😊 Adaptable
   😊 Powerful
   😊 Useful

😊 Tedious to configure and deploy
IMPROVEMENTS
Improvements: Configuration consolidation

- Overlap between configuration files

April 22, 2010
## Configuration consolidation (cont.)

<table>
<thead>
<tr>
<th>Name</th>
<th>Display Name</th>
<th>Description</th>
<th>Repeatable XPath</th>
<th>XPath Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>Title</td>
<td>This is the title of the image.</td>
<td>FALSE</td>
<td>mods:titleInfo/mods:title</td>
</tr>
<tr>
<td>dateTaken</td>
<td>Date Taken</td>
<td>Date the photograph was originally taken. Include as much detail as is available, including month and day when known. Enter in YYYY-MM-DD form.</td>
<td>FALSE</td>
<td>mods:originInfo/mods:dateCreated[encoding='W3cdtf' and @keyDate='yes']</td>
</tr>
<tr>
<td>description</td>
<td>Image Description</td>
<td>Provide a brief description of the content of the image.</td>
<td>FALSE</td>
<td>mods:titleInfo[@displayLabel='Image Description']/mods:title</td>
</tr>
<tr>
<td>sectors</td>
<td>Sectors</td>
<td>Appropriate Sector term(s) from DLC Sector vocabulary.</td>
<td>TRUE</td>
<td>mods:subject[@authority='dicsector']/mods:topic</td>
</tr>
<tr>
<td>regions</td>
<td>Regions</td>
<td>Appropriate Region term(s) from DLC Region vocabulary.</td>
<td>TRUE</td>
<td>mods:subject[@authority='dicregion']/mods:geographic</td>
</tr>
<tr>
<td>keywords</td>
<td>Keywords</td>
<td>Enter values from the DLC Subject Thesaurus. Do not use subdivisions.</td>
<td>TRUE</td>
<td>mods:subject[@authority='dickeyword']/mods:topic</td>
</tr>
<tr>
<td>photographer</td>
<td>Photographer</td>
<td>Enter the name of any known photographer in the form Last, First. If the photographer is not known, leave this field blank.</td>
<td>FALSE</td>
<td>mods:name[@type='personal']/mods:namePart</td>
</tr>
<tr>
<td>printLocation</td>
<td>Print Location</td>
<td>Select from the available values the location of the physical print or negative for this image. If the location is not known, leave this field blank.</td>
<td>FALSE</td>
<td>mods:location/mods:physicalLocation</td>
</tr>
<tr>
<td>previewId</td>
<td></td>
<td>FALSE</td>
<td>mods:location/mods:url[access='object in context']</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FALSE</td>
<td>mods:location/mods:url[access='preview']</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FALSE</td>
<td>mods:identifier[@type='local']</td>
<td></td>
</tr>
</tbody>
</table>
Example: Digital Library of the Commons

- Step one: Fill out spreadsheet
- Step two: generate configurations
- Step three: make any customizations
- Step four: catalog and browse
Improvements: Future Plans

- **Goals**
  - Be able to bring up new collections without:
    - A software developer/programmer
    - Waiting for a scheduled maintenance window

- **Improvements Needed**
  - Live/Hot configuration updates
  - Configuration interface
  - Bounds for configurability

- **Development Strategy**
  - Incremental improvements
    - Add functionality to template as needed
    - Generalize with every improvement
Thanks! Questions?