Using DSpace for Institutional Repositories

Heidi Butler, Kathryn Lightcap, and Stacy Nowicki
Great Lakes Information Technologists for Education and Research (GLITTER)
Kalamazoo College, 28 March 2008
Why an Institutional Repository?

- Digital access and storage of student theses
- Create a more accessible archive of college history and documents
- Access for faculty work
What’s This NITLE Thing?

- National Institute for Technology and Liberal Education
- Established September 2001
- Ithaka initiative with support from Mellon
- Collaborations among faculty, IT, and librarians
NITLE DSpace Pilot Beginnings

- LASR Project (Liberal Arts Scholarly Repository)
  - Introduced at Oberlin Group meeting, 2004

- Amherst, Carleton, Connecticut, Dickinson, Macalester, Middlebury, Simmons, Trinity

- ProQuest DigitalCommons™ software

- Wanted to expand the project – worked with NITLE
Meanwhile…

- Met with Beloit, Lake Forest, and Knox at the ACM office in June 2006
  - Content DM? ProQuest Digital Commons? DSpace?
  - Not sure if LASR would open up

- Asked NITLE for help, but NITLE already had LASR in the works

- LASR RFP answered by Longsight Group
  - Recommended DSpace
Dawn of the NITLE DSpace Pilot

- July 2006: NITLE pilot project open to Oberlin Group, CLAC, and NITLE member schools
- August 2006: NITLE takes over contractual obligations from LASR
- K signed up for DSpace pilot project in November 2006
- Implemented in January 2007 – LDAP authentication
Why DSpace? - 1

• Open Source
  – Cost Effective
  – Centralization of hosting/technical support
  – “Out of the Box” solution

• Sustainable
  – Commitments by major entities (MIT, HP, Mellon)
  – Currently in production

• Bonuses:
  – Crawled by Google and OA1ster!
  – Persistent URLs
Why DSpace? - 2

• Flexible enough for consortial use
  – Each college has its own branding
  – Multiple workflows, formats, metadata types
  – Consortium-wide searching
  – Authentication/permissions

• Control
  – Ability to define access policies at file (bitstream) level
  – Collections can have varying permissions
DSpace Structure

• Communities
  – Correspond to departments/offices/projects
  – Contains collections and sub-communities
  – Controlled by one or many
    • Ex: Department of Cute Puppies

• Collections
  – Types of data
    • Ex: Cute Puppy Honors Theses
Documents

• Policy

• Non-Exclusive Distribution License
  – Authors retain copyright, but K College can upload and distribute

• Metadata
  – Defines metadata elements
  – Required or optional
  – Communities, collections, items

• Instructions: uploading, cataloging, managing
K College Process - 1

1. Establish administrator
2. Define what the IR is, what it’s for
3. Write policies/“ideal” workflow plan
4. Involve representatives from all parts of the college (policies only)
5. Send policies around to group for input
6. Meet with group for suggestions
7. Write directions (“how to” materials)
8. Start practicing!
Eventually…

Marketing + Roll-out = Party time!
People Involved

- College Archivist
- Information Technology staff
- Cataloger
- Dean of Students
- Administrative Assistants
- Faculty
The Original Plan

- Use DSpace for access and preservation
- Uploading done by those who take responsibility for a “Community”
- Administrative Assistants responsible for student theses in each department
The Evolution

• Preservation not likely
  – Pilot project’s 100 GB down to 50 GB in June 2008
  – Too many big TIFF files for our storage capacity

• Most uploading done by IS staff and students

• Still learning how stuff works

• Contacting Administrative Assistants is a slow process
About the “K” Archives

- Staffed part-time + students only
- Have older but still OK tech setup
- Just beginning digital initiatives
- Collection includes approximately 23,000 images
- 175th anniversary = good time to take advantage of special funds 😊
What I’ve done in DSpace so far

• Test activities only
• Scanned, uploaded and sort of cataloged about 30 images
• Selected poster, several published volumes for others to scan and catalog
• Keep “mental file” of heavily used archival materials to eventually digitize
Good things

- Our DSpace peers are diverse
- Consortial effort (NITLE)
- Dublin Core metadata
- We’re in OAISter!
- Mapping of items (“bitstreams”)
- Thumbnail generation
- Subscriptions
I want us to...

- Establish more consistent standards for individual media types
- Sort out who is handling which aspects of the workflow process (or decide to just keep winging it)
- Find a repository for the stuff that won’t fit in DSpace (ie. TIFFs)

*Much of this will evolve with the outsourcing project later in 2008!*
My issues with DSpace...

- Not an intuitive system, even for tech-minded librarians
- The NITLE DSpace Help files are Not Very Helpful
- Dull public interface (is this our fault?)
- No controlled vocabulary feature in NITLE instance
- Still not sure what my role is, other than “tester”
Confusing examples

- “Authorizations”
- “Policies”
### File types accepted in DSpace

<table>
<thead>
<tr>
<th>Name</th>
<th>Extensions</th>
<th>MIME Type</th>
<th>Support Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td></td>
<td>application/octet-stream</td>
<td>unknown</td>
</tr>
<tr>
<td>Adobe PDF</td>
<td>pdf</td>
<td>application/pdf</td>
<td>known</td>
</tr>
<tr>
<td>XML</td>
<td>xml</td>
<td>text/xml</td>
<td>known</td>
</tr>
<tr>
<td>Text</td>
<td>txt, asc</td>
<td>text/plain</td>
<td>known</td>
</tr>
<tr>
<td>HTML</td>
<td>htm, html</td>
<td>text/html</td>
<td>known</td>
</tr>
<tr>
<td>CSS</td>
<td>css</td>
<td>text/css</td>
<td>known</td>
</tr>
<tr>
<td>Microsoft Word</td>
<td>doc</td>
<td>application/msword</td>
<td>known</td>
</tr>
<tr>
<td>Microsoft Powerpoint</td>
<td>ppt</td>
<td>application/vnd.ms-powerpoint</td>
<td>known</td>
</tr>
<tr>
<td>Microsoft Excel</td>
<td>xls</td>
<td>application/vnd.ms-excel</td>
<td>known</td>
</tr>
<tr>
<td>MARC</td>
<td></td>
<td>application/marc</td>
<td>known</td>
</tr>
<tr>
<td>JPEG</td>
<td>jpeg, jpg</td>
<td>image/jpeg</td>
<td>known</td>
</tr>
<tr>
<td>GIF</td>
<td>gif</td>
<td>image/gif</td>
<td>known</td>
</tr>
<tr>
<td>image/png</td>
<td>png</td>
<td>image/png</td>
<td>known</td>
</tr>
<tr>
<td>TIFF</td>
<td>tiff, tif</td>
<td>image/tiff</td>
<td>known</td>
</tr>
<tr>
<td>AIFF</td>
<td>aiff, aif, aifc</td>
<td>audio/x-aiff</td>
<td>known</td>
</tr>
<tr>
<td>audio/basic</td>
<td>au, snd</td>
<td>audio/basic</td>
<td>known</td>
</tr>
<tr>
<td>WAV</td>
<td>wav</td>
<td>audio/x-wav</td>
<td>known</td>
</tr>
<tr>
<td>MPEG</td>
<td>mpeg, mpg, mpe</td>
<td>video/mpeg</td>
<td>known</td>
</tr>
<tr>
<td>RTF</td>
<td>rtf</td>
<td>text/richtext</td>
<td>known</td>
</tr>
<tr>
<td>Microsoft Visio</td>
<td>vsd</td>
<td>application/vnd.visio</td>
<td>known</td>
</tr>
</tbody>
</table>

*continued…*
### File types accepted in DSpace

**continued...**

<table>
<thead>
<tr>
<th>Name</th>
<th>Extensions</th>
<th>MIME Type</th>
<th>Support Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMP3</td>
<td>fm</td>
<td>application/x-filemaker</td>
<td>known</td>
</tr>
<tr>
<td>BMP</td>
<td>bmp</td>
<td>image/x-ms-bmp</td>
<td>known</td>
</tr>
<tr>
<td>Photoshop</td>
<td>psd, pdd</td>
<td>application/x-photoshop</td>
<td>known</td>
</tr>
<tr>
<td>Postscript</td>
<td>ps, eps, ai</td>
<td>application/postscript</td>
<td>known</td>
</tr>
<tr>
<td>Video Quicktime</td>
<td>mov, qt</td>
<td>video/quicktime</td>
<td>known</td>
</tr>
<tr>
<td>MPEG Audio</td>
<td>mpa, abs, mpeg</td>
<td>audio/x-mpeg</td>
<td>known</td>
</tr>
<tr>
<td>Microsoft Project</td>
<td>mpp, mpd</td>
<td>application/vnd.ms-project</td>
<td>known</td>
</tr>
<tr>
<td>Mathematica</td>
<td>ma</td>
<td>application/mathematica</td>
<td>known</td>
</tr>
<tr>
<td>LateX</td>
<td>latex</td>
<td>application/x-latex</td>
<td>known</td>
</tr>
<tr>
<td>TeX</td>
<td>tex</td>
<td>application/x-tex</td>
<td>known</td>
</tr>
<tr>
<td>TeX dvi</td>
<td>dvi</td>
<td>application/x-dvi</td>
<td>known</td>
</tr>
<tr>
<td>SGML</td>
<td>sgml</td>
<td>application/sgml</td>
<td>known</td>
</tr>
<tr>
<td>WordPerfect</td>
<td>wpd</td>
<td>application/wordperfect5.1</td>
<td>known</td>
</tr>
<tr>
<td>RealAudio</td>
<td>ra, ram</td>
<td>audio/x-pn-realaudio</td>
<td>known</td>
</tr>
<tr>
<td>Photo CD</td>
<td>pcd</td>
<td>image/x-photo-cd</td>
<td>known</td>
</tr>
<tr>
<td>FMP3</td>
<td>fm</td>
<td>application/x-filemaker</td>
<td>known</td>
</tr>
<tr>
<td>BMP</td>
<td>bmp</td>
<td>image/x-ms-bmp</td>
<td>known</td>
</tr>
<tr>
<td>Photoshop</td>
<td>psd, pdd</td>
<td>application/x-photoshop</td>
<td>known</td>
</tr>
<tr>
<td>Postscript</td>
<td>ps, eps, ai</td>
<td>application/postscript</td>
<td>known</td>
</tr>
<tr>
<td>Video Quicktime</td>
<td>mov, qt</td>
<td>video/quicktime</td>
<td>known</td>
</tr>
</tbody>
</table>