How to Implement an Institutional Repository: Part II

A NASIG 2006 Pre-Conference
May 4, 2006

Technical Issues
Technical Issues

- Technical expertise
- Hardware and software
- Metadata support
- Interoperability
- Version control and revision
- User Interface
- Digital preservation
Technical expertise

- Knowledge of operating systems and servers
- Knowledge of database structure
- Ability to troubleshoot
Technical expertise

- Ability to install patches and updates
- Ability to pull together pieces from variety of sources
- Knowledge of standards
Hardware and software

- Open source
- Purchased or licensed
- Locally mounted or hosted externally
- Adequate server space
- Robust backup mechanisms
Software review

- Archimede
- ContentDM
- DSpace
- EPrints
- ETD-db
- Fedora
- Public Knowledge Project
Software requirements

- Permit the easy creation, use, and administration of digital objects distributed over the Internet
- Facilitate the creation of collections of materials in different disciplines or categories
Software requirements

- Support any type of file
- Carry out searches based on standard metadata
- Flexible metadata capture, edit, and display
Software requirements

- Plug into your local authentication system
- Be constructed using components and technologies that are standard and non-proprietary
- Easily integratable
Software requirements

- Customizable user interface
- Modular
- Flexible system administration
- Granular authorizations
Software requirements

- Scalable
- Manage licenses and permissions
- Recoverability
- Ease of managing underlying database
- Statistics and reports
Software requirements

- Flexible egress
- Flexible ingest
- Already implemented in other institutions
- Easy to set up
- Affordable
Metadata support

- Underlying metadata structure
- Ease of modification
- Global change capabilities
- Controlled lists of terms
Interoperability

What’s the objective?

- Facilitate sharing based on common standards
- Link digital archives around the world
- Provide access to metadata – and files
Interoperability

What’s the solution?

- OAI-PMH
  - Consistent interface
  - Minimal implementation
  - XML representation of Dublin Core metadata set
OAI Registries

Registry of Open Access Repositories (ROAR)

<table>
<thead>
<tr>
<th>Name</th>
<th>Summary Graph</th>
<th>Thumbnail</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) ask23 /archivsystem k23</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="http://ask23">http://ask23</a> /archivsystem k23 is an electronic, webbased platform for publishing and archiving digital media with a focus on artistic and scholarly work, publication and research. ask23 is a project of the Laboratory of Arts and Science, Academy of Fine Arts Hamburg, Germany. ask23 is eine Archiv- und Publikationsplattform für die künstlerische und wissenschaftliche Arbeit, zur Veröffentlichung von Textproduktionen und der Abfrage von digitalen Ressourcen sowie des analogen Archivs in Raum k23. Total OAI Records: 42 50% freely accessible fulltext (* estimate).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td><img src="image1.png" alt="Summary Graph" /></td>
</tr>
<tr>
<td>Other softwares (various) (OAI)</td>
<td></td>
<td><img src="image2.png" alt="Thumbnail" /></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) 11th Joint Symposium on Neural Computation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNU EPrints (OAI)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-Journal/Publication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>info:other:archives.eprints.org:import</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total OAI Records: 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100% freely accessible fulltext (* estimate)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
OAI Registries
Version control and revision

- Correct identification
- Security
- Revision
Public user interface

- Submission
- Searching
- Reuse of content
Ensuring the long-term maintenance of a bitstream (the zeros and ones):

- backing up files and keeping a copy at an offsite location
- running checks to track the deterioration of storage media, files or bitstreams
Digital preservation

- Providing continued accessibility of the contents:
  - viability
  - renderability
  - understandability
Digital preservation strategies

- Bitstream copying
- Refreshing
- Durable/persistent media
- Digital archaeology
- Analog backups
- Migration
- Emulation
Digital preservation components

- Metadata registry
- Format registry
- Checksum verification
- Backup procedures
- Persistent identifiers
## Metadata registry

### Dublin Core Type Registry

Note: Adding a new element to the DC Registry does not add a corresponding input field to the submit forms!

<table>
<thead>
<tr>
<th>ID</th>
<th>Element</th>
<th>Qualifier</th>
<th>Scope Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>contributor</td>
<td>advisor</td>
<td>Use primarily for thesis advisor.</td>
</tr>
<tr>
<td>3</td>
<td>contributor</td>
<td>author</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>contributor</td>
<td>editor</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>contributor</td>
<td>illustrator</td>
<td></td>
</tr>
</tbody>
</table>
### Bitstream Format Registry

**Extensions** are comma-separated lists of filename extensions used to automatically identify the formats uploaded files. Do not include the dot.

When you add a bitstream format, it is initially made "internal" so that it does not appear in the submission UI before you've finished editing the format metadata. Be sure to uncheck "internal" if the format should appear in the submission UI list of formats.

**More help...**

<table>
<thead>
<tr>
<th>ID</th>
<th>MIME Type</th>
<th>Name</th>
<th>Long Description</th>
<th>Support Level</th>
<th>Internal?</th>
<th>Extensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>application/octet-stream</td>
<td>Unknown</td>
<td>Unknown data format</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>text/plain</td>
<td>License</td>
<td>Item-specific license ag</td>
<td>Known</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>application/pdf</td>
<td>Adobe PDF</td>
<td>Adobe Portable Document</td>
<td>Known</td>
<td></td>
<td>pdf</td>
</tr>
<tr>
<td>4</td>
<td>text/xml</td>
<td>XML</td>
<td>Extensible Markup Language</td>
<td>Known</td>
<td></td>
<td>xml</td>
</tr>
<tr>
<td>5</td>
<td>text/plain</td>
<td>Text</td>
<td>Plain Text</td>
<td>Known</td>
<td></td>
<td>txt, asc</td>
</tr>
<tr>
<td>6</td>
<td>text/html</td>
<td>HTML</td>
<td>Hypertext Markup Language</td>
<td>Known</td>
<td></td>
<td>htm, html</td>
</tr>
<tr>
<td>7</td>
<td>application/msword</td>
<td>Microsoft Word</td>
<td>Microsoft Word</td>
<td>Known</td>
<td></td>
<td>doc</td>
</tr>
</tbody>
</table>
Checksum verification

Submit: Uploaded File

Here are the details of the file you have uploaded. Please check the details before going to the next step. (More Help...)

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>File Format</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barcelona_Marketing_Mundial.ppt</td>
<td>5082112 bytes</td>
<td>Microsoft Powerpoint (known)</td>
<td>2413ec89dd4b10c651a88176f6cfb4 (MD5)</td>
</tr>
</tbody>
</table>

Click here if this is the wrong format
Click here if this is the wrong file

You can verify that the file has been uploaded correctly by:

- Clicking on the filename above. This will download the file in a new browser window, so that you can check the contents.
- Comparing the checksum displayed above with a checksum worked out on your local computer. They should be exactly the same. Click here to find out how to do this.
Backup procedures

Preservation of Files in Scholars’ Bank

Scholars’ Bank accepts most file formats on deposit. File formats are tracked in a format registry.

As far as resources and technology permit, the University of Oregon Libraries will take steps to ensure that the files deposited in Scholars’ Bank survive uncorrupted and continue to be usable. The steps taken to ensure long-term viability of the materials include:

- nightly production of ASCII text backup, when possible, for those files that are text-based
- nightly incremental backup of files and metadata to the Libraries’ mass storage unit
- bi-weekly backup from the mass storage unit to tape
- bi-monthly export of the metadata and directory structure as a tar file to another system
- storage of magnetic tapes outside of the main library building
- conversion of some file formats upon deposit to more secure file formats
- periodic checking of file integrity to avoid data corruption
- monitoring the technological environment to prepare for file migration as file formats become obsolete or the software needed to run them becomes unavailable

Because the technological environment is changing rapidly, it may not be possible to preserve the usability of every file format that is deposited in Scholars’ Bank. The University of Oregon Libraries will make a good-faith effort to maintain the viability of the materials deposited in Scholars’ Bank or return them to their authors if unable to do so.

The University of Oregon Libraries are endeavoring to comply with the Open
Persistent identifiers

Title: 1st We Build Them, Then What? : The Future of Institutional Repositories

Other Titles: First We Build Them, Then What? : The Future of Institutional Repositories

Authors: Hixson, Carol G., 1955-

Keywords: Institutional repositories
Digital archives
Scholarly communication

Issue Date: 22-Nov-2005

Publisher: University of Oregon Libraries

Citation: BiD, textos universitaris de biblioteconomia i documentació; no. 15 (Jan. 2006)

Abstract: The article discusses the current state of IR development and outlines a direction that IRs in the U.S. can effectively take.

Description: 7 p. Published in January 2006 in BiD, textos universitaris de biblioteconomia i documentació.
Attributes of a Trusted Digital Repository

- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability
- System security
- Procedural accountability
- OAIS compliance
OAIS Reference Model
Test site