Institutional Repository using DSpace

Yatrik Patel
Scientist D (CS)
yatrik@inflibnet.ac.in
Institutional repositories [are] digital collections capturing and preserving the intellectual output of a single or multi-university community.

A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.

It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.
Institutional Repositories are.

Centred around a university (or academic institution) and contain items which are the **scholarly output of that institution**

- A **collection of (digital) objects**, in a variety of formats
- Include works of **various degrees of scholarly authority** and from various stages in the process of scholarly inquiry.
- In addition to published works, an IR may include preprints, theses & dissertations, images, data sets, working papers, course materials, or anything else a contributor deposits
- Typically motivated by a commitment to **open access**
In India, Government Funds for approximately 75% of education and 95% of research.

The results of research funded by government should ideally be accessible in open access to the society at large.

Articles in open access have greater impact in terms of citation rates than those published in subscription-based journals.

Funding agencies like NIH in US & Welcome Trust in UK have already mandated submission of results of research funded by them in open access journals or make them available in open access.
Open Access: Publishing Cycle
Institutional Repositories

- Are organized around a particular institutional community
- Often are dependent upon the voluntary contribution of materials by scholars for the content in their collection
- Are mainly repositories and therefore may only offer limited user services

Digital Libraries

- May be built around any number of organizing principles (often topic, subject, or discipline)
- Are the product of a deliberate collection development policy
- Typically include an important service aspect (reference and research assistance, interpretive content, or special resources.)
IR : Services

Institutional Repositories
(Services and Related Works)

- Organization of digital information
- Information retrieval
- User interface
- Archiving and preservation
- Services and social issues
- Evaluation and applications to specific areas
Starting IR

1. Justify the relevance to the institution and contributors
2. Develop a policy framework. How will we find this content and what will we do with it?
3. Build the infrastructure
4. Bonus: Get institutional support and a mandate.
IR – Key Issues

- Faculty buy-in
- Submission policies
- Intellectual Property issues
- Mediated deposit
- Metadata
- OAI-PMH compliant systems
- Specialized staff
- Outreach and Liaison services
## Expectations from Institutional Repository Solution

<table>
<thead>
<tr>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Effective (Hardware, Software and Maintenance)</td>
</tr>
<tr>
<td>Technically simple to install and manage</td>
</tr>
<tr>
<td>Robust</td>
</tr>
<tr>
<td>Scalable</td>
</tr>
<tr>
<td>Open and inter-operable</td>
</tr>
<tr>
<td>Modular</td>
</tr>
<tr>
<td>User Friendly</td>
</tr>
<tr>
<td>Multi-user (Both Aspects User and Administration)</td>
</tr>
<tr>
<td>Platform independent</td>
</tr>
<tr>
<td>Capable of handling multimedia digital objects</td>
</tr>
</tbody>
</table>
Widely Used Systems

Produced by Berkeley Electronic Press (bepress), focused on maintaining scholarly output. Not open source.

Developed at the University of Southampton (UK). Widely considered to be the least complex of the major repository software platforms.

Developed at Cornell and University of Virginia. Based on a framework known as the Flexible Extensible Digital Object and Repository Framework.

Designed by MIT and Hewlett-Packard to manage the intellectual output of research institutions and provide for long-term preservation.
What is DSpace

DSpace is a platform that

• capture items in any format – in text, video, audio, and data. It
• distributes it over the web. It indexes digital items, so users can
• search and retrieve items.
• preserves digital content over the long term.

DSpace is typically used as an institutional repository or digital library. It has three main roles:

• Facilitate the capture and ingest of materials, including metadata about the materials
• Facilitate easy access to the materials, both by listing and searching
• Facilitate the long term preservation of the materials

DSpace is a joint project of MIT Libraries and Hewlett-Packard Labs. Now being handled by DuraSpace
The beginning: 2000

- The DSpace project was initiated in July 2000 as part of the HP-MIT alliance.

Software releases:

- Version 1.0 – 8th November 2002
- Version 1.1 - 8th May 2003
- Version 1.2 – 13th August 2004
- Version 1.3 – 3rd August 2005
- Version 1.4 – 26th July 2006
- Version 1.5 – 25th March 2008
- Version 1.6 – 2nd March 2010
- Version 1.7 - 17th December 2010 (End of Life JAN 2014)
- Version 1.8 - 4th November 2011
- Version 3.0 - 30th November 2012
- Version 3.1 - 30 Jan 2013
- Version 3.2 -24 July 2013
- Version 4.0 - 16 December 2013
Key Factors to DSpace’s adoption

- Open source, freely available
- Greate support network of current users World Wide
- Easy to use as packaged
- Can handle a multitude of digital formats
- Initially developed by leading institutions
- Content all accessible through Google Scholar
DSpace can be used to store any type of digital medium.

- Journal papers
- Data sets
- Electronic theses
- Reports
- Conference posters
- Videos
- Images
- Almost everything....
Open source software (www.dspace.org)
- BSD licence

Community development model
- Source code control repository (SVN)
- Committers
- Community welcome to submit bug reports, patches, feature requests
- Email lists for support
The DSpace Information Model
DSpace Information Model

Communities
• Research units of the organization

Collections (in communities)
• Distinct groupings of like items

Items (in collections)
• Logical content objects
• Receive persistent identifier

Bitstreams (in items)
• Individual files
• Receive preservation treatment
Communities & Collections

Collections and Communities organize items into a hierarchical form.

Metadata:
- Limited descriptive metadata available
- Name, description, license, etc...

Example:

Communities and Collections
- **Institutional Documents and Records**
  - Annual reports
  - Bills and other Documents
- **Scientific contributions**
  - Agricultural Sciences
  - Anthropology
  - Biological Sciences
  - Computer Science
  - Economics

Communities

Collections
Items

- Items are logical units of content

Metadata:
- All items have qualified Dublin Core metadata
- May contain metadata in other formats encoded as a bitstream

Example:
- E thesis
- Book
- Web page (Images, CSS, HTML)
- Photographs
Item Metadata

Descriptive

- Qualified Dublin Core
- Non Dublin Core is also supported
- Any other format may be added as a bitstream
  - However, it will not be searchable

Administrative

- Who can access, remove, or modify an item
- Stored in the database, no standard format used

Structural

- Very basic
- What bitstreams are contained in an item
- What collections and communities does an item belong to
Bitstreams are Individual Digital files

Metadata:
- Limited descriptive metadata available
  - name, file format, size, etc...

Example:
- PDF file
- Word document
- JPEG picture
- Executable program
- HTML file
- CSS file
Bundles

Bundles group related bitstreams together

Metadata:
• No metadata

Example:
• HTML files and images that compose a single HTML document may be organized into a bundle
• Typical bundles are:
  — ORIGINAL
  — THUMBNAILS
  — TEXT
  — LICENSE
  — CC_LICENSE
Components & Features of DSpace
Metadata registry

* Maintain what metadata fields may exist for an item in DSpace.
* Three components
  * Schema
  * Element
  * Qualifier
  * Scope Note
Maintain a registry of file formats

Three levels:
- Supported
- Known
- Unknown
DSpace user accounts are called E-people

If permitted, an e-person may:

- Login to the site
- Sign up to receive notifications about changes to a collection
- Submit new items to collections
- Administer collections/communities
- Administer the DSpace site.
Authorization

The DSpace authorization system enables administrators to give e-people the ability to perform the following operations on an object.

- **Add / Remove**
  - Enable an e-person to add or remove any object (community, collection, item)
- **Collection Administrator**
  - Enable an e-person to edit an item’s metadata, withdraw items, or map items into the collection.
- **Write**
  - Enable an e-person to add or remove bitstreams
- **Read**
  - Enable an e-person to read bitstreams
Ingestion = getting contents into DSpace

Batch import
- Many at a time
- Needs to be in a specific format
  - XML encoded metadata
  - Bitstreams

Web based submission
- One at a time
- Workflow processes
Workflow

* Step 1: May reject the submission
* Step 2: Edit metadata or reject
* Step 3: Edit Metadata

Source: www.dspace.org (Dspace Documentation)
Search & Browse

Users may browse any item in DSpace

- Title
- Author
- Date
- Community / Collection
- Subject

Users may search for any item in DSpace based upon any Dublin Core value or a full text search.
Handle System

Provides a persistent identifier

Standard URL’s change

- Hardware or software changes
- Political changes
- Network changes

Handles attempt to address these problems by creating a permanent URL independent of the repository.

Example:

- http://hdl.handle.net/1944/225
OAI-PMH

Enables other sites to harvest metadata from a DSpace repository

Collections are exposed as OAI sets

Only Dublin Core metadata is available
Statistics

Analyses the DSpace logs to generate a set of statistics on how DSpace is being used.

Metrics collected:

- Number of items archived
- Number of bitstream views
- Number of item page views
- Number of collection page views
- Number of community page views
- Number of user logins
- Number of searches performed
- Number of license rejections
- Number of OAI Requests

Presented in a by-month form or in-total form.
SWORD and OpenURL Support

SWORD (Simple Web-service Offering Repository Deposit) is a protocol that allows the remote deposit of items into repositories. DSpace supports the SWORD protocol via the 'sword' web application and SWord v2 via the swordv2 web application.

DSpace supports the OpenURL protocol from SFX. If your institution has an SFX server, DSpace will display an OpenURL link on every item page, automatically using the Dublin Core metadata.
Areas one can customize

- Submission process: one can configure the submission steps to suit organization
- Browse and search terms: can set what fields and files you choose to index and display in the browse interface
- Database: can choose Postgres or Oracle
- Extend DSpace to work with other web services: using Light Network Interface you can pull or push content to/from DSpace
- User interface: you can create your own user interface
Welcome to ShodhGanga@INFLIBNET Centre

The ShodhGanga@INFLIBNET Centre provides a platform for research students to deposit their Ph.D. theses and make it available to the entire scholarly community in open access. The repository has the ability to capture, index, store, disseminate and preserve ETDs submitted by the researchers.

Theses and dissertations are known to be the rich and unique source of information, often the only source for research work that does not find its way into various publication channels. Doctoral dissertations are manifestation of result of four to five years of intense work involving huge investment of resources, both mental and physical and infrastructure and other support from the universities. A thesis reflects quality of research work conducted by a student and the ability of an institution to lead and support original work of research in a given discipline.

As per the UGC Regulation on M.Phil/Ph.D., the responsibility of hosting, maintaining and making the digital repository of Indian Electronic Theses and Dissertation (ShodhGanga) accessible to all institutions and universities is assigned to the INFLIBNET Centre.

Universities Contributed in ShodhGanga [First 40 Universities]

Choose a University to Browse its Departments. (3350+ theses uploaded)

[ Show All Universities ]

A. A. Alagappa University [2]  
Avinashilingam Deemed University for Women [7]  
Amrita Vishwa Vidyaapeetham (University) [19]  
Banaras Hindu University [1]
Welcome to the Digital Repository of the Raman Research Institute

This contains the research publications of the faculty and students of the Raman Research Institute. The collected papers of C.V. Raman and the historical records of the Institute (Annual Reports and Newspaper Clippings) are also housed here.

Search

Enter some text in the box below to search DSpace.

Communities in DSpace

Choose a community to browse its collections.

01. Archives [752]
02. C.V. Raman and his work [516]
03. Science at the Raman Institute: 1948-1970 [52]
04. Astronomy and Astrophysics [1272]
05. Light and Matter Physics [275]
06. Soft Condensed Matter [1182]
07. Theoretical Physics [740]
08. Library & Information Services [20]