

OPEN SOURCE SOFTWARE INSTITUTION REPOSITORY: A STUDY WITH SPECIAL REFERENCE TO DSPACE USERS IN INDIA

M. B. Khatri*, Vaishali G. Choukhande**

Abstract *Old technologies of information exchange are being replaced by new technologies and methods. In this age Internet has become an essential medium for information exchange and with its helps to maintain scholarly communication anyone can communicate his/her information with another at anywhere in the world. The result is that from last few years open access, digital publishing material and the related activities of self archiving and creating institutional repositories have taken place of old methods. It's a new technique for digital college building, managing preserving information and creating new information in digital form in institutional and about institution or college. By using this repository the institution can offer services like dissemination of information, access to preserve and use information and as well as content submission and organization of information. The present paper focuses on the Institutional Repositories, its characteristics, advantages and disadvantages, availability of different types of software's, Dspace one of the best software building Institutional Repositories, supports from Dspace, analysis of Dspace, its users from India and the information provided by repository.*

Keyword: *Institutional Repositories, DSpace, Open source Software, Self Archiving, Scholarly communication.*

INTRODUCTION

In this age Internet has become an essential medium for information exchange and it helps to maintain scholarly communication anyone can communicate his/her information with another at anywhere in the world. The result is that from last few years open access, digital publishing material and the related activities of self archiving and creating institutional repositories have taken step. It's a new technique for digital college building, managing preserving information and creating new information in digital form in institutional and about institution or college. By using this repository the institution can offer services like dissemination of information, access to preserve and use information and as well as content submission and organization of information.

INSTITUTIONAL REPOSITORY

Repository means formally organized and managed collection of digital content generated by faculty, staff and students at an institution. The main purposes of institutional repositories are to bring together and preserve the intellectual output of a laboratory, department and commitment.

Wikipedia an encyclopedia described IR as "An institutional Repository is an online locus for collecting and preserving in digital form, the intellectual output of an institution particularly a research institution". For a university, this world include materials such as research journal articles

before and after undergoing peer review and digital versions of theses and dissertations but it might also include other digital assets generated by normal academic life, such as administrative documents, courses notes or learning objects. Open access is one step ahead of free access. It holds to remove all barriers whether it is the matter of prizes or permission by using scientific communication medium and internet and Institutional Repository is a step towards open access. *According to Lynch* "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. In a university setting an Institutional Repository may provide a place for faculty work, students theses and dissertations, e-journals, datasets and so on."

SOFTWARE AVAILABLE

Numbers of software's are available in the market for institutional Repository. Some of them are on pay basis and some are free downloaded software for ex.: DSpace, GNU Eprints CDsware, Fedora, Greenstone etc. In this software DSpace, Greenstone, Eprints are free software.

Dspace

DSpace is the software of choice for academic, non-profit, and

* Librarian, S.R.C., Ayurved College, Chikhli, Dist. Buldana (M.S.) India. Email: ajaykhatri13@gmil.com

** HOD, LIS, Sant Gadge Baba Amravati University, Amravati, Email: vchoukhande@rediffmail.com

Table 1. Analysis of Dspace Features

Feature	Dspace
URL	http://www.dspace.org
Software Availability	Free & more reputed
Operating System required	Unix or Linux
Language used	Java
Technical support	http://www.dspace.org/feedback/mailing.html
License	BSD License
System Required Language	Apache We server, Tomcat, servile engine & Postgre, SQL relational database, Java 1.4 or latter
Training	Available
Developers	Massachusetts Institute of Technology Library (MIT) & Hewlett Packard company
Security	Fool proof security system
Current Version	1.8.2

commercial organizations building open digital repositories. It is free and easy to install “out of the box” and completely customizable to fit the needs of any organization. DSpace preserves and enables easy and open access to all types of digital content including text, images, moving images, mpegs and data sets. And with an ever-growing community of developers, committed to continuously expanding and improving the software, each DSpace installation benefits from the next.

Dspace is very popular software for developing open source Digital library. The MIT (Massachusetts Institute of Technology) Library and Hawlett-Packard Laboratory has been jointly developed it. Under the BSD open license for research institution Dspace is available for capturing, indexing, storing, preserving and redistributing data. Dspace is an important software. It is a perfect solution for institutional repository to manage and serve and reuse intellectual output of an institution. Dspace is a digital asset management system and creation, indexing searching storing and retrieval can possible. Dspace is available free basis, so that the needy can use it easily.

Dspace support following items: Computer program, Published books, Administrative records, Articles, preprints, technical papers, Theses/ dissertations, Images, Audio, video, Web pages, Multimedia publication

Technical support: Dspace-tech is a media by which one can help by Dspace developers. They share needed information, helps other and solve technical problems. Dspace- wiki has been used by Dspace developers for keeping in touch with and for sharing code files. Unix or Linux operating system are used to run Dspace. Dspace can be easily installed and maintained by developers.

Document format: Dspace supports different file formats text, audio, video & images. The plug-in is used to ingest document, like PDF, XML, HTML, RTF, Plain text, Excel, PPT, etc.

OAI support: Dspace is OAI PMH compliant. It defines the metadata standards based on Dublin core for the purpose of interoperability. It uses OAI PMH through OAI cat tool for harvesting metadata. Dspace is default uses qualified Dublin core set for finishing metadata.

Institutional Repository: Dspace is one of the best and easier software for building institutional repository. In Dspace different forms of scholarly publication are archived and preserve.

Features of Dspace : Digital creation management and distribution, Security can be build at various levels to affect access, Long term physical storage & management of digital material, Indian language based libraries can be built. It conforms to the UNICODE standards Librarian and researcher can upload & download digital items for all over the world.

Why use DSpace

- Largest community of users and developers worldwide
- Free open source software
- Completely customizable to fit your needs
- Used by educational, government, private and commercial institutions
- Can be installed out of the box
- Can manage and preserve all types of digital content

Current Version

DSpace 1.8.2 can be downloaded from the files area in Source Forge (Zip) or via SVN. DSpace 1.8.2 is a bug fix release for the 1.8.x platform, which resolves issues discovered in either 1.8.1 or 1.8.0.

Features and Improvements

DSpace 1.8 is a major release of DSpace. It contains many new features, numerous bug fixes and system improvements, including:

- Improvements to the upgrade and configuration process: Most notably dspace.cfg has been split up, and the default behavior when doing 'Ant update' is to replace the existing files if the new files are different.
- More Curation Tools/Plugins: Including a new "Metadata Link Checker" and "Metadata Translator".
- Configurable Workflow for the XMLUI: Allows administrators to customize their workflow approval process.
- SWORD Client: Enables DSpace to push/submit content to other SWORD enabled repositories.
- SWORDv2 Server Module: DSpace can now support submissions via SWORD protocol version 2 (as well as version 1).
- RSS feeds now support richer features, such as iTunes podcast and publishing to iTunesU.
- Enhancements to Discovery faceted searching & browsing system.
- Rewrite of Creative Commons licensing for XMLUI.
- Ability to now reorder bit streams in an item.
- Ability to enable automatic virus checking during item submission.
- Ability to Withdraw/Reinstate/Delete Items in Bulk, via Batch Metadata Editing.

ANALYSIS AND INTERPRETATION OF DATA

Classification Users as Per Type of Institute Using Dspace

Table no.1- Highlight the classification of the users by their institute types. It is find that most of the institute using the DSpace are academic institutes the percentage of them is near about 74%. None of the national library is using the same for any purpose. As the same none of the museum, consortium, medical center Hospital, consulting services has been accepted this software. The maximum institutes are

academic institutes after those government and researcher centers are important users of this software they are each 9%. 2% are the Non profitable and Personal users. The commercial users of this software from India are only 3%.

Table 1. Classification users as per type of Institute using DSpace

Sr No.	Types of Institutes	Number
1	Academic	65
2	Government	08
3	Non Profit	02
4	Research Center	08
5	Museum Gallery	00
6	National Library	00
7	Commercial	03
8	Personal	02
9	Consortium	00
10	Medical Center Hospital	00
11	Consulting Services	00
12	Other	01

User Classification by Product Version

Table no.2 focuses the software has different version as updates from their stats. Today there are ten updated versions are available for the users. The mostly used version is 1.5 S JSP which is used by 27% users. The Version no. 1.4 X is also one of the most used version in India the 19% of the Indian users are still using this version. The version no 1.8 X JSP has no users in India. The most current version 1.8 X JSP XML Manakin has just a single user in India.

Table 2: User Classification by Product Version

Sr No.	Version	Number
1	1.3 X	08
2	1.4 X	12
3	1.5 X JSP	17
4	1.5 X JSP XML Manakin	06
5	1.6 X JSP	07
6	1.6 X JSP XML Manakin	06
7	1.7 X JSP XML Manakin	03
8	1.7 X JSP	04
9	1.8 X JSP	00
10	1.8 X JSP XML Manakin	01

Classification of Users by Database Used

To establish any database user has to use a backhand server. Table no.3 shows the different types of databases used by the Indian users of the DSpace. 65% of the users are using PostgreSQL for their database. Oracle is the used by only 4% users. Ms SQL has 12% users and 18% users are using different types of databases.

Table 3. Classification of Users by Database Used

Sr No.	Database Type	Number
1	PostgreSQL	32
2	Oracle	02
3	Ms SQL	06
4	Other	09

Classification of Users by Repository Access Authority

Once the Database is created, the question arises that who should be the user to that. The table no.4 classified the user by the repository access authorities. Most of the repositories are established for the public access. The percentage of them is 73%, 16% repositories are for both public and non-public access. 13% repositories are only for non-public user's means only for institute related people.

Table 4. Classification of Users by Repository Access Authority

Sr. No.	Access	Number
1	Public	56
2	Non- Public	10
3	Both	13

Classification of Users by Case Used

Table no.5 focuses on the cases used by the institute while establishing their repository. The software is used by different repository cases in India. The software is popular in India for establishing institutional repository, 28% users are using for institutional repository, 20% users are using for subject repository, 18% users are using this software for creating learning resources, 10% users are from audio video repository and image repository and 5% and 3% user are each from government records reports and others.

Table 5. Classification of users by Case used

Sr No.	Cases	Number
1	Institutional Repository	33
2	Audio Video Repository	12
3	Image Repository	12
4	Govt. Records Reports	06
5	Federated or Networked	00
6	Subject Repository	24
7	Museum Cultural	01
8	Learning Resources	21
9	Other	04

Classification of users as per types of Files included

Once institute start to create repository they has to decide which types of data files they should attach. Table no.6 describes the different types used by the different types of the Indian users. 35% of the files from India are text document files which are reposted with the help of DSpace. The 18% users used image files, 13% users used audio files, 11% users used video files, 10% users used website links, 9% users used datasets in their repository.

Table 6. Classification of Users as Per Types of Files Included

Sr No.	File Type	Number
1	Audio	23
2	Video	20
3	Image	32
4	Text Documents	64
5	Data Sets	17
6	Website	18
7	Other	07

Classification of Users by Content Type

Table no. 7 describes the content type and it's used. 23% of the contents in DSpace repository are Conference Papers and Presentations. 17% contents are Research or Development Organization Project, 11 % are reference documents. Very few repositories has contents for Technical reports or work papers, community cultural, history art archives, Infrastructure or Admin Records, Special Subject, Infrastructure or Admin Records.

Table 7. Classification of Users by Content Type

Sr. No.	Content	Number
1	Faculty Student Output	00
2	Learning Resources	07
3	History Art Archives	08
4	Community Cultural	02
5	Research or Development Organization Project	27
6	Government Documents	10
7	Reference Documents	17
8	Tech Reports or Work Papers	06
9	Conference Papers and Presentations	37
10	Data Sets	13
11	Special Subject	09
12	E-mails	10
13	Infrastructure or Admin Records	02
14	Metadata Only	09
15	Other	02

13	Lightweight JHOVE Format Validator and Virus Check	02
14	Tombstoning	01
15	Google Analytics	00
16	Google Indexing	03
17	Language Packages	06
18	Manakin Themes	06
19	OAI Harvester Plugin	00
20	SWORD	03
21	Integration of Sakai	06
22	Integration of Open Journal System	08
23	Integration of DOME	06
24	Integration of Moodle	03
25	Integration of Shibboleth	04
26	Creative Commons Open URL	08
27	SFX	03
28	Other	01
29	Mirage	01
30	DSpace Discovery	05

Classification of Users by Integrations or Customizations

Table no. 8 highlights the classification of users by integration or customization of the repository. To customization of the repository 11% of the institutes used Dublin Core Meta Toolkit, 9% users used Statistics for better use and integration. Controlled Vocabulary Ontology, Pluggable Storage S3, Recommendation, Request Copy, Lightweight JHOVE Format Validator and Virus Check, Tomb stoning, Google Indexing are some of the frequently used tools.

Table 8. Classification of users by Integrations or Customizations

Sr. No.	Integration or Customization	Number
1	Commenting	07
2	Dublin Core Meta Toolkit	14
3	Embargo	05
4	Joomla Collections Archives and Repositories	06
5	Lightweight Network Interface	04
6	Controlled Vocabulary Ontology	02
7	Pluggable Storage S3	02
8	Recommendation	02
9	Request Copy	05
10	Semantic Search	00
11	Statistics	11
12	Web of Communications	04

FINDINGS

- DSpace is the very popular Open source software in India, used for Institutional repository.
- Most of the users of this software in India are Academic institutes.
- 1.5 X JSP and 1.4 X are the most used versions in India.
- The very much used database for the repository by Indian institutes is PostgreSQL.
- Almost repositories are established for public use.
- Text documents as well as audio, video files, datasets, web links are repudiated in India.
- Conference Papers and Presentations are commonly used content in repository.

CONCLUSION

Library cultural is turning towards digitization. Today most of the libraries are automated. Libraries have always been engaged in managing and developing their collection. Libraries have long experience for collection building and maintain the information. Institutional Repository and open access will be the next challenge for Library and LIS professionals. The Libraries have long experience with developing and managing content collections. LIS professional are always in front of new technology and they accept it and trying to develop their library services. DSpace preserves and enables easy and open access to all types of digital content including text, images, moving images,

mpegs and data sets, the features in DSpace is more than any other software.

REERENCES:

- Das, Anup Kumar (Ed.), Open Access to Knowledge and Information, UNESCO, 2008, ISBN 978-81-89218-21-8, en.wikipedia.org/wiki/Institutional_repository (accessed on 10/09/2012)
- Ghosh, S. B., Open access and Institutional Repository a developing country perspective: A Case Study of India, 72nd IFLA General Conference 20-24 Aug 2006
- Rao, P. Venkata, Institutional Repositories: A Key Role For Libraries, International CALIBER 2007-Panjab University, Chandighar. pp 689-695
- Robinson, Mary, Institutional Repositories: Staff and Skills requirements, 8th August 2007 university of Nottingham, <http://www.sherpa.ac.uk/documents/sherpaplusdocs/nottsRepository%20Staff%20and%20Skills.pdf> (accessed on 10/09/2008)
- Singh, S. N. and other, Open Source Software: A comparative Study of Greenstone and Dspace, International CALIBER 2007-Panjab University, Chandighar. Pp 89-97
- Vishala, B.K., and Bhandi, M.K., Building Institutional Repository (IR): Role of the Library, International CALIBER 2007-Panjab University, Chandighar. pp 631-640
- Maureen Pennock, Stuart Lewis, M. Pennock, S. Lewis Institutional Repositories: The New University Challenge Article for ALISS Quarterly, April 2007 http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/aliss_rsp.pdf