academicJournals

Vol. Vol. 5(10), pp. 313-318, November, 2013
DOI: 10.5897/IJLIS2013.0322
ISSN 2141–2537 ©2013 Academic Journals
http://www.academicjournals.org/IJLIS

International Journal of Library and Information Science

Review

Open source software: An institutional digital repository system with special reference to DSPACE software in digital libraries - an introduction

VELMURUGAN C.

SIVA Institute of Frontier Technology, India.

Accepted 5 July 2013

The institutional repositories are powerful systems that allow institutions to store and maintain their digital documents and permit interaction and collaboration among users in the organizations. There is a number of digital library software available as "Open Source" as well as in "Proprietary format". Open source software helps libraries mainly in lowering initial and on-going costs, eliminating vendor lock-in and allowing greater flexibility. The main advantage of open source software is that it is generally available in free. DSpace is a ground breaking digital library system to capture, store, index, preserve and redistribute all scholarly research material in digital formats. This paper explains the open source software with special reference to DSpace software. It describes the DSpace system, types, content and data models, highlights the technical architecture of DSpace and finally highlights some comments about the future development and operation of the DSpace digital repository system.

Key words: Open access, open source software, DSpace, digital library, preservation, institutional repository, digital repository, open archives.

INTRODUCTION

The history of open source software began with the early stages of computer and software development. At that, time programmers and developers frequently shared their software freely. Advent of companies in software development with the aim of profit making restricted the culture of sharing source code of software. Milestones in the history of open source Software are:

1. 1983 - Richard Stallman formed GNU project.
2. 1985 – Creation of Free Software Foundation.
3. 1991 – Development of Linux kernel by Linus Torvalds.
4. 1998 – Open Source Initiative (OSI) formed by Eric Raymond.

The two terms "free" and "open source" have been used synonymous for free distribution of softwares. Popular licenses used for this purpose are the GNU General Public License (GPL), BSD license, GNU Lesser General Public License, MIT License, Mozilla Public License and Apache License. All these licenses have some differences in their terms and conditions; they ensure users' freedom, copying, distribution and improvement of software. Fundamentals of these licenses are similar to the philosophy of Free Software Foundation. "Free software is a matter of the users' freedom to run, copy, distribute, study, change and improve software" (Kumar, 2008).

The Open Source Software for libraries portal (http://www.oss4lib.org), established in mid 1999, listed some of the library related projects in Table 1. These range from simple scripts to statistics production, integrated library systems and institutional repository software.

E-mail: murugan73@gmail.com.

Table 1. Types of open source software.

Name	URL	Type of project
Apache	http://www.apache.org	Web server
Free BSD	http://www.freebasd.org	Unix Operating System
GIMP	http://www.gimp.org	OS image manipulation Software
GNOME	http://www.gnome.org	Unix desktop environment
KDE	http://www.kde.org	Unix desktop environment
Linux	http://www.linux.org	Unix Operating System
Mozilla	http://www.mozilla.org	Web browser
My SQL	http://www.mysql.org	Database
Project Gutenberg	http://promo.net/pg/	Freely available digital content (Started 1971)
Open Office	http://www.openoffice.org	Office application suite
PHP	http://www.php.net	OS Programming tool
DSpace	http://www.dspace.org	Digital Library Software
Eprints	http//www.eprints.org	Digital Library Software
Greenstone	http://www.greenstone.org	Digital Library Software

DSPACE DIGITAL REPOSITORY SYSTEM:

Dspace is a digital repository system that captures, stores, indexes, distributes and preserves an organization's research data. D-space is the software of choice for academics, non-profit and commercial organization building open digital repositories. It is free and easy to install and completely customizable to fit the needs of any organizations. D-space is jointly developed by MIT libraries and Hewlett-Packard labs released in April 2004. D-space integrates a user community orientation into a system's structure. This design supports the participation of the schools, departments, research centers and other units typical of a large research institution. D-space was developed in response to expressed faculty need for an easy to use, dependable service that could manage, host, preserve and distribute materials in any type of digital medium formats, that is, Journal papers, Data sets, Electronic theses, Reports, Conference posters, Videos, Images, jpeg, mpeg, tiff files.

RELATED STUDIES

Bertot and McClure (1998) also evaluated nine open access repositories in the field of Computer Science and Information Technology. The repositories have been evaluated using content; preservation policies; right management; promotion advertisements; services; feedback and access status as important parameters. Fernandez (2006) reflects the status of open access repositories across India. Carpenter et al. (2011) have also envisioned new features in the institutional repository world. Workflow pattern in institutional repositories has been researched by Hanlon and Ramirez (2011). A shifting landscape of institutional repositories is well knitted by various authors (Shreeves and Cragin, 2008; Nykanen, 2011). DSpace is not a completely new concept because 'Preprint' archives in the sciences were already in existence as a form of innovation managed by the data creators (Smith, 2003). According to Westell (2006), DSpace and the other technical solutions that came to prominence around the same time are 'institutional' in scope. Individual institutions have accepted ownership rights and responsibility for management and preservation of their local scholarly corpus. Universities also want "to address the issues of repatriating their scholarly work from commercial publishers and providing long term, secure and open access". Since the launch of DSpace, institutional repositories have sprung up at academic institutions across the world and more are appearing regularly. Repository management has been well researched by number of authorities (Bide, 2002; Genoni, 2004; Medeiros, 2003; Poynder, 2006; Markey et al., 2007; McDowell, 2007). Metadata issues in institutional repositories have been researched by Dunsire (2008) and Goldsmith and Knudson (2006). Lynch (2003) has also discussed the infrastructure of institutional repositories visualizing the future developments also.

DSPACE architecture

There are three layer architecture based on DSpace namely, (a) application layer (b) business layer (c) storage layer (Figure 1).

Application layer: The application layer covers the interface to the systems, the web and user and interface and batch loader, in particular.

Business layer: The business layer is where the DSpace-specific functionality resides, Including the

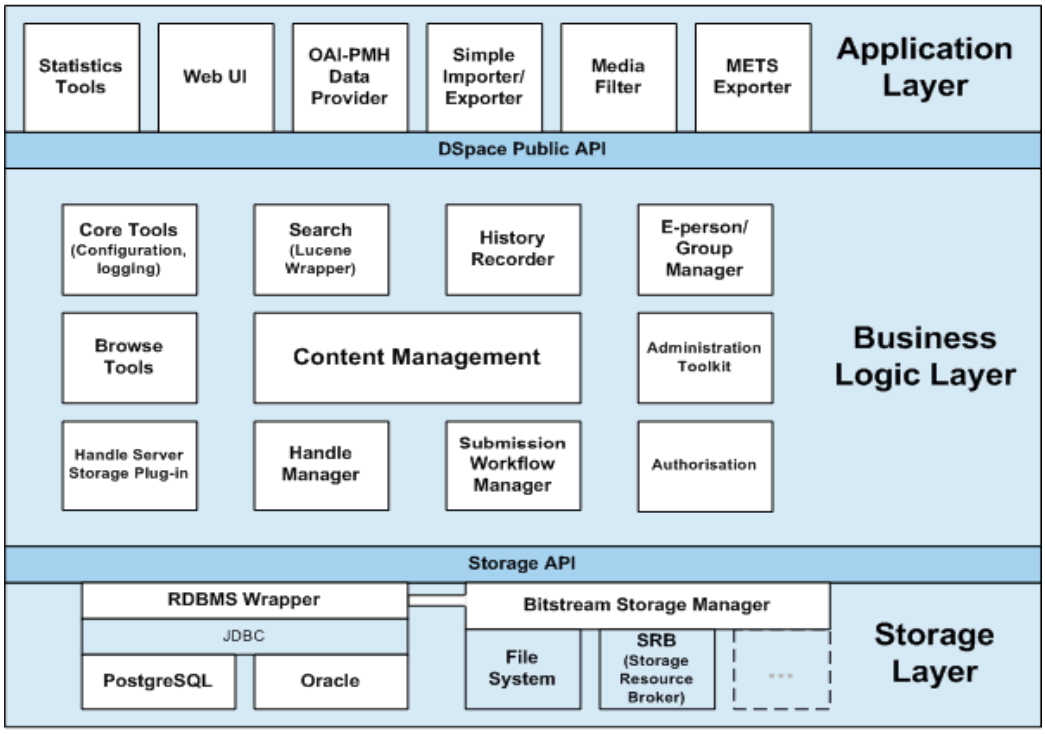

Figure 1. Types of DSpace architecture. Source: http://www.dlib.org/dlib/january03/smith/fig-2.gif.

workflow, content management, administration, and search and browse modules. Each module has an API to allow Dspace adopters to replace or enhance that function as desired.

Storage layer: The storage layer is implemented using the file system, as managed by PostgreSQL database tables.

The system is primarily written in Java, and uses only free software libraries and tools, including the Postgre-SQL, RDBMS, Java servlet, Apache and tomcat, Lucene search engines, XML tools and RDF tool. Collections within communities consist of items. Items are, in turn, composed of one or more bit streams, or physical files of digital materials. DSpace item is single bit stream, for example of a digital image encoded as a TIFF file, or a digital documents encoded as a PDF file. A final example is a digital document that consists of a set of several HTML pages and some JPG images.

REASONS TO USE DSPACE

The most important reasons to use the Dspace open source software are as follows:

1. Free open source software
2. Register of reliable commercial service partners
3. Largest community of users and developers worldwide
4. Completely customizable to fit your needs
5. Granular, collection based access and submission rights
6. Compliant with embargo and licensing policies on restricted documents
7. Manages and preserves all types of digital content: text, images, moving images, presentations, datasets
8. Turnkey institutional repository application, delivering a working solution out of the box.
9. Optimal Metadata management for submission by non-technical end users
10. Customizable import procedures for virtually any structured metadata source
11. Standardized export and harvesting features supporting integration and migration
12. Optimal web page structure for information retrieval in Search Engines.
13. Creates persistent network identifiers (citeable URLs) for content.
14. Indexes content so end users can easily browse, search and retrieve information.

DSpace can be used to store any type of digital materials, including:

1. Documents, such as articles, preprints, working papers, technical reports, conference papers
2. Books
3. Theses

Table 2. Notable examples of repository other than India with URL.

No	Software	Institute	URL
1	Dspace	BRAC University Institutional Repository	http://dspace.bracu.ac.bd/
2	Dspace	International Centre for Diarrhoeal Disease Research Digital Repository, Bangladesh (ICDDR,B)	http://dspace.icddrb.org/
3	Dspace	Xiamen University Institutional Repository, China	http://dspace.xmu.edu.cn/
4	Dspace	DSTO Scientific Publications Online Repository, Australia	http://dspace.dsto.defence.gov.au/dspace/
5	Dspace	University of the West Indies at Mona , Jamaica -	http://www.mona.uwi.edu/
6	Dspace	Instituto Tecnologico de Costa Rica	http://www.tec.cr/
7	Dspace	Bolivarium, Venezuela	http://dspace.bolivarium.usb.ve/dspace/

4. Data sets
5. Computer programs
6. Visualizations, simulations, and other models
7. Multimedia publications
8. Administrative records
9. Published books
10. Overlay journals
11. Bibliographic datasets
12. Images
13. Audio files
14. Video files
15. e-formatted digital library collections
16. Learning objects
17. Web pages

Notable institutional repository other than India

These are seen in Table 2.

DSPACE DIAGRAM- DESCRIPTION

1. Web-based interface makes it easy for a submitter to create an archival item by depositing files. DSpace was designed to handle any format from simple text documents to datasets and digital video.
2. Data files, also called bitstreams, are organized together into related sets. Each bitstream has a technical format and other technical information.
3. An item is an "archival atom" consisting of grouped, related content and associated descriptions (metadata). An item's exposed metadata is indexed for browsing and searching. Items are organized into collections of logically-related material.
4. A community is the highest level of the DSpace content hierarchy. They correspond to parts of the organization such as departments, labs, research centers or schools.
5. DSpace's modular architecture allows creation of large, multi-disciplinary repositories that ultimately can be expanded across institutional boundaries.
6. DSpace is committed to going beyond reliable file preservation to offer functional preservation where files are kept accessible as technology formats, media, and paradigms evolve over time for as many types of files as possible.
7. The end-user interface supports browsing and searching the archives. Once an item is located, Web-native formatted files can be displayed in a Web browser while other formats can be downloaded and opened with a suitable application program.

DATA MODEL DIAGRAM

The way data are organized in DSpace is intended to reflect the structure of the organization using the DSpace system. Each DSpace site is divided into *communities*, which can be further divided into *sub-communities* reflecting the typical university structure of college, department, research center, or laboratory.

Communities contain *collections*, which are groupings of related content. A collection may appear in more than one community. Each collection is composed of *items*, which are the basic archival elements of the archive. Each item is owned by one collection. Additionally, an item may appear in additional collections; however every item has one and only one owning collection. Items are further subdivided into named *bundles* of *bitstreams*. Bitstreams are, as the name suggests, streams of bits, usually ordinary computer files. The Bitstreams that are somehow closely related, for example HTML files and images that compose a single HTML document are organized into bundles (Table 3).

The following actions are possible:

Note that there is no 'DELETE' action. In order to 'delete' an object (an item) from the archive, one must have REMOVE permission on all objects (in this case, collection) that contain it. The 'orphaned' item is automatically deleted (Table 4).

DSPACE DIRECTORIES

A complete DSpace installation consists of three

Table 3. Objects in the DSpace data model.

Object	Example
Community	Laboratory of Computer Science; Oceanographic Research Center
Collection	LCS Technical Reports; ORC Statistical Data Sets
Item	A technical report; a data set with accompanying description; a video recording of a lecture
Bundle	A group of HTML and image bitstreams making up an HTML document
Bitstream	A single HTML file; a single image file; a source code file
Bitstream	Format Microsoft Word version 6.0; JPEG encoded image format

Table 4. Objects with actions in the DSpace data model.

Object	Action
Community	
ADD/REMOVE	add or remove collections or sub-communities
Collection	
ADD/REMOVE	add or remove items (ADD = permission to submit items)
DEFAULT_ITEM_READ	inherited as READ by all submitted items
DEFAULT_BITSTREAM_READ	Inherited as READ by Bitstreams of all submitted items. Note: only affects Bitstreams of an item at the time it is initially submitted. If a Bitstream is added later, it does not get the same default read policy.
COLLECTION_ADMIN	Collection admins can edit items in a collection, withdraw items, map other items into this collection.
Item	
ADD/REMOVE	add or remove bundles
READ	can view item (item metadata is always viewable)
WRITE +	can modify item
Bundle	
ADD/REMOVE	add or remove bitstreams to a bundle
Bitstream	
READ	view bitstream
WRITE	modify bitstream

separate directory trees. It is important to get a general understanding of the DSpace directories and the names by which they are generally referred (Figure 2).

1. **Installation directory,** referred to as *[dspace],* is the location where DSpace is installed and running off of it is the location that gets defined in the dspace.cfg as "dspace.dir". It is where all the DSpace configuration files, command line scripts, documentation and webapps are installed to.
2. **Source directory,** referred to as *[dspace-source]:* This is the location where the DSpace release distribution is unzipped into. It usually has the name of the archive that you expand such as *dspace-<version>-release* or *dspace-<version>-src-release*. It is the directory where all of your "build" commands is run.
3. **Web deployment directory:** This is the directory that contains your DSpace web application(s). In DSpace 1.5.x and above, this corresponds to *[dspace]/webapps* by default. However, if you are using Tomcat, you may decide to copy your DSpace web applications from *[dspace]/webapps/* to [tomcat]/webapps/ (with [tomcat] being wherever you installed Tomcat--also known as $CATALINA_HOME).

Conclusion

Open source primarily offers useful savings in terms of time, money, and resources. To save and preserve library data for future, it is important that libraries adopt as many OSS as they can and participate in the movement of sharing information globally with open standards and open formats. Dspace is the most popular among the digital library solutions available in the open source domain. The Dspace is very powerful digital

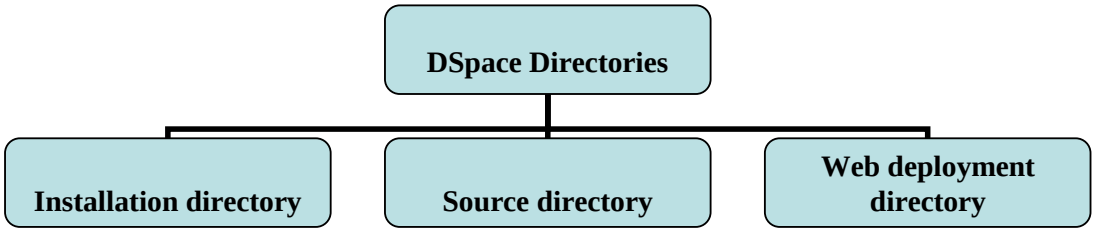

Figure 2. Three DSpace directory trees.

library software. The major advantage of the software is that it allows submission of digital documents by its members. Educational institutions dominate in the use of these packages. However, many institutions have implemented digital libraries, but not all are online. The knowledge of Open access is possible only if these repositories are only made online. We anticipate getting more institutions involved, and to working together with them to achieve real goal and change.

REFERENCES

Bertot JC, McClure CR (1998). Measuring electronic services in public libraries; issues and recommendations. Public Libraries 37(3):176-180

Bide M (2002). Open archives and intellectual property: incompatible world views? Open Access Forum, Bath. Retrieved from www.oaforum.org/otherfiles/oaf_d42_cser1_bide.pdf.

Carpenter M, Graybill J, Offord Jr. J, Piorun M (2011). Envisioning the Library's Role in Scholarly Communication in the Year 2025. Portal: Libraries Acad. 11(2):659-681. doi:10.1353/pla.2011.0014

Dunsire G (2008). Collecting metadata from institutional repositories. OCLC Systems & Services: International digital library perspectives. 24(1):51-58. doi: 10.1108/10650750810847251.

Hanlon A, Ramirez M (2011). Asking for Permission: A Survey of Copyright Workflows for Institutional Repositories. portal: Libraries Acad. 11(2):683–702. doi: 10.1353/pla.2011.0015

Fernandez L (2006). Open access initiatives in India: An evaluation. The Canadian Journal of Library and Information Practice and Research, 1(1). Retrieved from: http://www.dlib.org/dlib/january05/foster/01foster.html.

Genoni P (2004). Content in institutional repositories: a collection management issue. Library Manage. 25(6-7):300-306. doi: 10.1108/01435120410547968.

Goldsmith B, Knudson F (2006). Repository librarian and the next crusade: the search for a common standard for digital repository metadata. D-lib Magazine 12(9). Retrieved from http://dlib.ukoln.ac.uk/dlib/september06/goldsmith/09goldsmith.html

Kumar V (2008). Selection and management of open source software in libraries. Asian School of Business, Padmanabha Building, Techno Park, Trivandrum. Retrieved May 6.5.2013, from http://eprints.rclis.org/archive/00008739/01/OSS-selection management.pdf.

Lynch C A (2003). Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. Portal: Libraries Acad. 3(2):327-336. doi: 10.1353/pla.2003.0039

Markey K, Rieh SY, St. Jean B, Kim J, Yakel E (2007). Census of Institutional Repositories in the United States: MIRACLE Project Research Findings. Washington, D.C: CLIR. Retrieved from http://www.clir.org/pubs/reports/pub140/pub140.pdf

McDowell CS (2007). Evaluating institutional repository deployment in American academe since early 2005: Repositories by the numbers, Part 2. D-Lib Magazine 13(9/10). Retrieved from http://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html.

Medeiros N (2003). E-prints, institutional archives, and metadata: disseminating scholarly literature to the masses. OCLC Systems Serv. 19(2):51-53. doi: 10.1108/10650750310481757.

Nykanen M (2011). Institutional Repositories at Small Institutions in America: Some Current Trends. J. Electronic Resourc. Librarianship. 23(1):1-19. doi: 10.1080/1941126X.2011.551089.

Poynder R (2006). Clear blue water. Retrieved from http://poynder.blogspot.com/2006/03/institutional-repositories-and-little.html.

Shreeves SL, Cragin MH (2008). Introduction: Institutional Repositories: Current State and Future. Library Trends 57(2):89-97. doi: 10.1353/lib.0.0037.

Smith M (2003). "An open Source Dynamic Digital Repository "Retrieved from: http://www.dlib.org/dlib/january03/smith/01smith.html

Westell M (2006). Institutional repositories: proposed indicators of success. Library Hi Tech. 24(2):211-226.