Open Source Solutions for creation of ETD archives/repository:
A case study of Central Library @ NIT Rourkela

D. P. Tripathi[1]
tripathidp@nitrkl.ac.in

Bhojaraju Gunjal[2]
gunjalb@nitrkl.ac.in

D. K. Pradhan[1]
dkpradhan@nitrkl.ac.in

[1]Assistant Librarian and [2]Head & Deputy Librarian,
Central Library, National Institute of Technology Rourkela, Odisha, 769008. INDIA.

Abstract

In academic/research institutions, the Institutional Repositories (IR) is the primary source to archive scholarly communication containing data, information, photographs and results of valuable research. These are vital and robust systems that allow institutions to preserve and maintain their digital documents and allow for interaction and collaboration among users in the different institutions. There are numbers of digital library software available as "Open Source" as well as "Proprietary". The open source software helps the libraries especially in eliminating vendor lock-in and reducing the cost with greater flexibility. These software are generally available free of cost with the capability to capture, store, index, preserve and redistribute all scholarly research materials in digital formats.

This paper deals with open source software (OSS) solutions especially DSpace, EPrints, Greenstone, Fedora projects and others for establishing institutional repository (IR). These OSS have developed a system that allows users to use and submit the contents for IR, and also allows the administrators to organize, configure and customize the IR accordingly to the requirements. The primary attempt of this paper is to discuss the implementation of IR with special reference to Odiya Books & Electronic Theses and Dissertations (ETDs) in Central Library of National Institute of Technology Rourkela (NITR). An attempt has been made to showcase the implementations of various OSS tools and usage ranking of these websites with the help of screenshots.

Keywords

Open Source Software, Electronic Theses and Dissertations, DSpace, Greenstone, EPrints, Fedora

1. Introduction

Libraries in the present era are facing problem of insufficient budget for experimentation & implementation of technology and also procurement of resources. To manage the information resources, there is a necessity of software and tools. This situation is demanding the libraries to stand up alone with the latest technology. Most of the software being used in these libraries are 'proprietary' which means there is involvement of money without the actual source code availability. On the other hand, Open Source Software is just opposite which revolves around sharing and collaboration. By nature, OSS is free for anyone to copy, modify and redistribute which increases the possibility of being used as per the requirements. There are number of open source software available globally which could be used in library such as basic Operating Systems, Integrated Library Management System (ILMS), Digital Library Software and Document Processing Systems etc [1].

The services and the collection of the libraries and information centers are on high demand due to the application of information and communication technology. Due to shrinking budgets and the increasing prices of resources, librarians have to look forward for alternative solution by which they can collect, store, arrange, and disseminate information to the users. The concept of open access and institutional repository (IR) has evolved to find out the solutions [2].

2. **Open Source Software concept**

The Open Source Software is a programme refers to two different related things [2]:

- **Source code:** A collection of instruction written in human readable computer language.

- **Object code:** A sequence of instructions in a computer language, usually a machine code language i.e. binary.

3. **Scope of this paper**

The scope of this paper is (i) to compare various OSS for building digital repository and (ii) to find out user-friendly OSS based on popularity for creation of digital repository for the prospective institute.

4. **Methodology**

For this paper, both observation method for data collection and case study methods are followed with reference to Central Library of NIT Rourkela. The following steps are followed for data collection by [2]:

- Surfing internet for IR open source software details

- Evaluating of the software for different perspectives

- Selecting the appropriate software based on different criteria

5. **Open Source Software: An Overview**

There are many open source software available for creating Institutional Repository but only 10 OSS have been selected for the study purpose based on its popularity.

5.1 DSpace (http://www.dspace.org)

DSpace is open source digital repository software which is used for archiving digital content. It mainly focuses on the long-term storage, easy accessibility and procurement of digital content. This software is developed by Hewlett Packard and MIT, USA released under a GPL license. The DSpace is compiled with Java and utility programs for maintaining assetstore and associated metadata store. It provides the interface for system administration to deposit, modify, search and access the content. DSpace supports PostgreSQL and Oracle relational database management system for storing metadata, accessibility and configuration with two web interfaces JSPUI and XMLUI. It also supports OAI-PMH v2.0 to export METS (Metadata Encoding and Transmission Standard). The latest version of DSpace supports faceted search and browsing functionality using Apache Solr [7].

5.2 E-Prints (http://www.eprints.org/uk/)

EPrints is open source software used for creating digital repositories compliant with open archives initiative protocol for metadata harvesting with Perl based plugin architecture for importing and exporting data and converting objects (for search engine indexing). The EPrints software is developed by University of Southampton School of Electronics and Computer Science released under GPL license. It is based on LAMP architecture written in Perl and PHP. It supports the Windows, Linux and Mac OS X. The overall appearance of a repository can be controlled using HTML templates, CSS and image. This software is available in many languages [8].

5.3 Greenstone (http://greenstone.org)

Greenstone produced by New Zealand Digital Library Project at the University of Waikato and developed and distributed in cooperation with UNESCO is open source software for creating and distributing collection of digital library. It supports Multilanguage and is released under the terms of General Public License.

The main objective of this software is to empower users especially in university libraries and other public service oriented institutions for building their own digital libraries. The complete interface of Greenstone and documentation is available in many languages such as French, Spanish, Russian, English etc [9].

5.4 Fedora Project (http://fedora-commons.org/)

Fedora (Flexible Extensible Digital Object Repository Architecture) is open source repository software for the management and dissemination of digital content for access and preservation. It can be used for complex and very large digital collection such historic and cultural materials as well as scientific data. It is capable of archiving different types of content and its metadata with various storage options. The information can be accessed via web API and web based administrator GUI. It also supports & provides RDF search, OAI-PMH, GSearch (Fulltext) and rebuild utilities with multiple & customer driven front-ends [10].

5.5 Invenio (http://invenio-software.org/)

Invenio is also open source software developed by the CERN Document Server Software Consortium to manage digital assets of any organization or institute. Mainly this software is used for managing scholarly or published digital content. This software is capable of organizing the collection in navigable collection tree having powerful search engine with customizable simple and advanced search interface, combined metadata, Fulltext and citation search in one go. It also supports flexible metadata standard (MARC) for handling articles, books, theses, photos, videos and many more with customizable display and linking rules. It also produces the output in multi format such as MARC, XML, HTML and OAI [11].

5.6 Omeka (http://www.omeka.org/)

Omeka is open source content management system to manage the digital collection for building archives in any institute or organization to publish and exhibit cultural heritage objects. It is very lightweight in nature in comparison to other repository software such as DSpace, Greenstone, and Fedora etc. It supports Dublin Core Metadata to display the contents. This software is developed by the Roy Rosenzweig Center for History and New Media at George Mason University and being used by many renowned libraries such as New York Public Library, Newberry Library as well as many small libraries [12].

5.7 DAITSS (http://daitss.fcla.edu/)

DAITSS is open source digital repository software developed by the Florida Centre for Library Automation. This software is capable of repository management which enables submission, Ingest, Archiving, Accessing, and withdrawal. It supports RESTful web services and micro services for archived content and also implements active procurement strategies. This software is very well suited for materials available in the form of text, documents, images, and audio-video formats [13].

5.8 Dienst (http://www.cs.cornell.edu/cdlrg/dienst/software/DienstSoftware.htm)

Dienst is open source software for managing the services running on distributed servers to cooperate in providing the digital library services. This software is supported by the Computer Science Technical Reports Project, a DARPA-funded collaboration to establish a digital library of computer science technical reports [14].

This software can be used in three ways:

- Institutions / Organizations which are willing to join an existing digital library built using Dienst.
- Institutions / Organizations which are willing to create a new distributed digital library with Dienst.
- Institutions / Organizations which are willing to undertake research in digital libraries using an existing Dienst digital library or by experimenting with the software.

5.9 ETD-db (http://scholar.lib.vt.edu/ETD-db/index.shtml)

The ETD-db is open source software written using Perl Script with MySQL database developed by Virginia Tech. This software provides a standard interface for web users, researchers and library professionals to manage the various fields and metadata related to collection of ETD.

The software is currently maintained by Paul Mather, and is available free of charge to members of the NDLTD [15].

5.10 IR-Plus (http://code.google.com/p/irplus/)

IR+ is open source software for building institutional repository and managing its collection with easy features to understand and manage. The main aim of this software is to meet the requirement of any organization/institutions to publish and preserve digital information / content [16].

The major features of the IR+ are:

- Group work spaces - allows a group of people to create a common area to share folders and files.
- Group project pages - similar to researcher pages but will allow a group of users to highlight a group project and expose files in a group workspace.
- Auto-share personal folder - allow a user to set automatic sharing of a file added to a folder.

6. Comparative study of OSS features

Some of the popular OSS have been selected for the comparative study w.r.t. various perspectives. The selection criteria of these OSS have been made based on the wide usage of OSS around the world. Table 1 mentioned below shows the current software available for creating the ETD/Archive with URL to download. These software are freely downloadable and can be used for the said purpose.

Table 1: Comparative study of OSS features

Attributes	DSpace	E-Prints	Fedora	Greenstone	ETD-db
Platform	Window & Linux	Windows & Linux	Linux	Windows, Linux, MAC OS	UNIX-based
Server	Apache Tomcat	Apache	Tomcat	Apache	Apache
Supporting Software	Java, Maven, Ant	PHP, Perl	Java, Jetty	Java, Image Magic	
Database	PostgreSQL & Oracle	MySQL	Many storage options (database and file systems)	Flat file	MySQL
Language	Java	PHP based	Java	C++, Perl	Perl
Scalability	Stable & can be upgraded	Stable & can be upgraded	Stable & can be upgraded	Stable & can be upgraded	Stable & can be upgraded
Standards Compliance	OAI-PMH	OAI-PMH	OAI-PMH Provider Service	Metadata extraction	OAI-PMH & XML encoded MARC
Search	OpenSearch	Open Search	GSearch (Fulltext) Search Service	Automatic Indexing	Open Search
Report generation	Yes	Yes	Yes	Yes	Yes

7. Open Source Repository Software @ NIT Rourkela: a case study

On 15th August 1961, the institute was founded as the Regional Engineering College, Rourkela (REC) under joint management of the Government of India and Government of Orissa for imparting quality education in the field of Engineering and Technology. The college was designated, on 26th June 2002, as the National Institute of Technology, a fully autonomous Institute supported by the Ministry of Human Resource Development, Government of India. On 15th August, 2007, on its 47th Foundation day, the Institute was declared to be an Institute of National Importance by an act of the Parliament. Over the years, NIT Rourkela has strived to establish itself as one of the premier national level institutions for carrying out front line research and imparting quality technical education in the country [17].

Today, NIT Rourkela is a highly prestigious institute with a reputation for excellence in research, consultancy and education at undergraduate, postgraduate and doctoral levels. It is passionately committed to make our country a world leader in technology and science and to inculcate this commitment among all its students. The target is to be known around the world for the academic standards and to be counted among the best technological institutes of India in terms of innovation, entrepreneurship and creation of intellectual wealth [17].

For archiving and disseminating the scholarly output, the central library of NIT Rourkela has been using the following software.

7.1 DSpace for Institutional Repository (http://dspace.nitrkl.ac.in/dspace)

The Central Library of NIT Rourkela has been using DSpace since 2006 for managing Institutional Repository for the scholarly contents of the institute which includes Book Chapters, Journal Articles and Conference Proceedings.

The Central Library has recently upgraded DSpace (from version 1.2 to 4.2) with the following practical features:

- Authentication & Authorization
- Dynamic HTML support
- OAI-PMH Support & Object Management
- Import / Export, Report Statistics
- Handle System
- Customization & Types of document supported
- Standards Compliance, Optimized Search and Browse
- Ranking, Embargo

It was really a tough task to upgrade DSpace from version 1.2 to 4.2 as there had been a long gap in release of different version of DSpace. It had the problem of compatibility of database PostgreSQL as version 7 was used in 1.2 and 9.2 in DSpace 4.2. Availability of proper documentation was another challenge to proceed with. The dependency on DSpace forum was a great help to start with hit and trail to up-grade and reach up-to the version DSpace 4.2. At last, the rigorous effort made successful up-gradation of DSpace.

Figure 1. DSpace version 1.2

Figure 2. DSpace version 4.2

7.2 E-Print for Electronic Theses & Dissertations (http://ethesis.nitrkl.ac.in/)

The central library has two other types of collections i.e. Electronic Theses & Dissertations and Odiya Books which were to be archived in order to provide the better access to these resources. The BPCL evaluated many open source software based on the following criteria in order to select the suitable software for the purpose. These are as follows:

- Authentication & Authorization
- Dynamic HTML support
- Object Management
- Import / Export
- General Statistics
- Customization & different types of document supported
- Standards Compliance
- Optimized Search and Browse
- Embargo

The installation of E-Prints was done with the version 3.1 in the year 2006. Later, up-grading E-Prints was tough task as it was to be changed from one database structure to the next database structure. In order to upgrade the E-Prints, the major help was received from the forum and also from the different institutes who have already upgraded their repositories to next version.

Figure 3. E-Prints version 3.1

Figure 4. E-Prints version 3.3

7.3 E-Print for Open Access Odiya Book Project (OAOB) (http://oaob.nitrkl.ac.in/)

OAOB is a not-for-profit project, initiated by National Institute of Technology, Rourkela, Srujanika, Bhubaneswar and Pragati Utkal Sangh, Rourkela. It disseminates the available information in Oriya language treasured in our local libraries and individual collections. The project is made available online to millions of readers worldwide. An open access repository containing resources pertaining to Oriya literature and culture, OAOB provides free access to researchers, scholars, historians, librarians, and the general public at all point in time.

The mission of the project is to preserve cultural history of Orissa by digitizing old, rare and new literary and other texts written in the Oriya language.
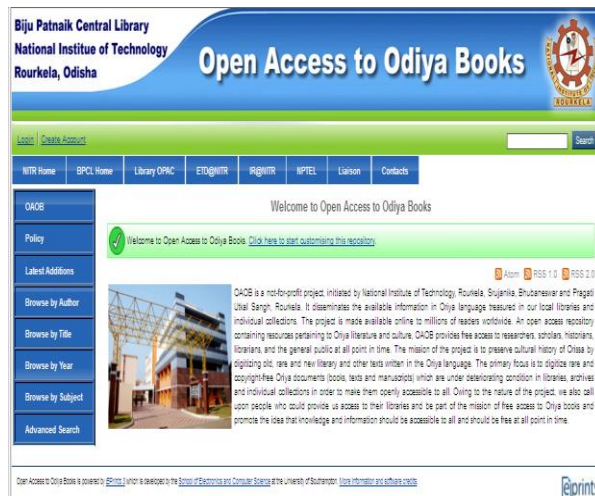

Figure 5. E-Prints version 3.1

Figure 6. E-Prints version 3.3

7.4 Greenstone Digital Library software:

In the early bid of time, the NIT Rourkela started its digital library project with GSDL to archive the ETD and subject question papers. In later stage, this GSDL software helps in organizing the proprietary software

procured by the Institute with its demonstration. Presently, this software helps the central library to organize the procured CD-ROM database in ISO format.
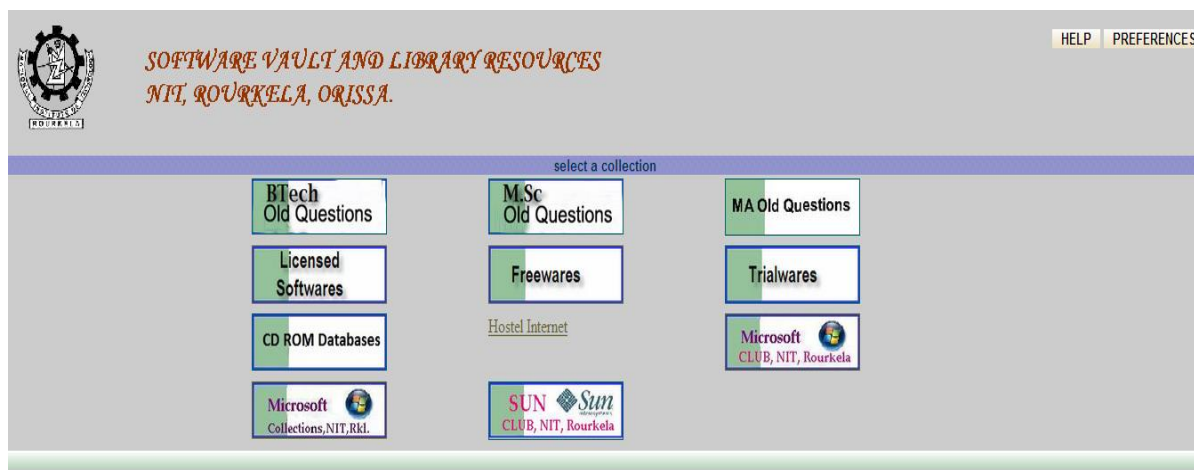

Figure 7. Greenstone DL for various resources

8. **Challenges for adopting OSS in libraries:**

Setting up Infrastructure for Archives

- **Hardware**
    - Linux Server with CentOS
    - Fiber Optic Network (Connects Academic Department, Hostels, Library and Other Centers)
- **Software**
    - **DSpace & EPrints for IR**
        - Multi Format support
        - Customizable Interface
        - OAI-PMH Complaint
        - Fully Qualified DCMI with facility of remote submission
        - Authorization & Reviewing
        - Collection Architecture (Community & Sub-community)
        - Import & Export Features
        - Persistent Identifier (Handle Number)
        - Search Engine
    - **Image Editing & OCR**
        - Abbyy Fine Reader
        - Fireworks
- **Configuration and Customization**
    - E-mail Settings (Admin, Submission, Error handling, feedback)
    - Digital Content Storage
    - Boolean Search Operator
    - Fields to Index for Search
    - Handle Settings
    - Embargo Settings
    - Item Exports and Download Settings
    - Settings for Item Count Information
    - Upload File Settings (Size and Directory)
    - Statistical Reports Configuration Settings

- **Selection of Metadata Standard**
  - To describe the content and attribute of any particular item and also facilitate searching and discovery

Designing Standard Workflow Pattern

Uploading mode of submission
  - CD and Online Submission (proposed)

Digitization Process:

- Acquiring & Selection and Duplicate Check
- Determining Copyright
- Preparing the Object
- Digital Conversion
  - Scanning,
  - Text Conversion,
  - Format,
  - Cleansing & OCR
  - File Naming
- Metadata Preparation
- Digital Archiving
- Providing Access

Figure – 8

Some of the challenges to be taken care of in libraries for adopting OSS are as follows:

- Archival Policies
- Proper user training of library staff
- Support from technical expertise
- Frequent up-gradation of the OSS
- Proper database creation and up-dation
- Enabling OSS to comply with aforesaid attributes

9. **Conclusion**

Due to tremendous growth in inter and multi-disciplinary subjects, there has also been growth in research outputs. The main challenge for every academic library is to manage these resources in effective way to serve the academic fraternity and provide the research output. The details mentioned above in the paper describe the different available open source software for creating ETD/Archive in order to manage the research output effectively. Though, every software has the capability to meet the requirements of **IR** and **ETD** but **DSpace, E-Prints and Greenstone** have more popularity due to its simplicity, flexibility, easy to use and easy accessibility through web-browser. This particular paper mainly focuses on the available Open Source Software for creating Archive/ETD with special reference to central library, NIT Rourkela.

References:

1. Randhawa, Sukhwinder (2012). "Open Source Software for Creation of Digital Library: A Comparative Study of Greenstone Digital Library Software & DSpace" Indian Journal of Library and Information Science Volume 6 Number 3(Supl), Sept - Dec 2012

2. Lihitkar, Shalini R. and Lihitkar, Ramdas S. (2012). "Open Source Software for Developing Digital Library: Comparative Study" DESIDOC Journal of Library & Information Technology, Vol. 32, No. 5, September 2012, pp. 393-400

3. Surendran Cherukodan, G. Santhosh Kumar and S. Humayoon Kabir, (2013),"Using open source software for digital libraries", The Electronic Library, Vol. 31 Iss 2 pp. 217 – 225

4. Biswas, Gautam and Paul, Dibyendu (2010). "An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library" International Journal of Library and Information Science Vol. 2(1) pp. 001-010, February, 2010

5. Tramboo, Shahkar [et.al.] (2012). "A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone" International Journal of Computer Applications (0975 – 8887) Volume 59– No.16, December 2012

6. Naik, Umesha and Shivalingaiah, D, "Digital library open source software: a comparative study" from http://ir.inflibnet.ac.in/bitstream/1944/529/1/4%28cal%2006%29.pdf

7. DSpace http://www.dspace.org

8. EPrints http://www.eprints.org/uk/

9. Greenstone http://www.greenstone.org/

10. Fedora http://fedora-commons.org/

11. Invenio http://invenio-software.org/

12. Omeka http://omeka.org/

13. DAITSS (http://daitss.fcla.edu/)

14. Dienst http://www.cs.cornell.edu/cdlrg/dienst/software/DienstSoftware.htm

15. ETD-db http://scholar.lib.vt.edu/ETD-db/index.shtml

16. IR-Plus http://code.google.com/p/irplus/

17. NITR http://nitrkl.ac.in