

Application of Open source solutions for digital asset management: a case study of NIT Rourkela

By

Vinod Kumar Mishra
Assistant Librarian
NIT Rourkela
Email: mishravk79@gmail.com

Puspita Mishra
Assistant Librarian
NIT Rourkela
Email: ppanda2010@gmail.com

Abstract

This paper attempt to highlight some of the most popular open source solutions used in Biju Patnaik Central Library (BPCL) of National Institute of Technology Rourkela (NITR), Odisa, India for digitization, preservation and dissemination of digital assets. The management and technical challenges faced in the organization and dissemination of such digital resources has been also described. The paper discusses about the various open source solutions like EPrint, DSpace, Joomla, Koha etc. used for digital preservation and dissemination of knowledge worldwide. It also describes some of the technical and policy related issues faced in the entire management of digital resources and their probable solutions.

Keywords: DSpace, Eprint, Joomla, Koha, Digitization, Automation, Greenstone

1. Introduction

Digital libraries are complex information systems. With the arrival of computers, the concept centered on large bibliographic databases, the now familiar online retrieval and public access systems. After forming the internet, the concept evolved again and turned to creating digital libraries. Phrases like, ‘virtual library’, ‘electronic library’, ‘library without walls’ and most recently ‘digital library’ are used time to time.

Digital Libraries have received much attention, allowing access to digital information from anywhere across the world. They have become widely accepted and even preferred information sources in different fields of study. The rapid growth of Internet and the increasing interest in the development of digital library related technologies and

collections helped to accelerate the Digitization of printed documents in the past few years.

Libraries are now more focusing on collection development of electronic resources like e-journals, e-books, e-databases, CD/DVD etc. as these resources gives more satisfaction to users and they can carry/transfer them very easily in electronic devices like, mobile, laptop etc. At the same time libraries are also facing challenges for proper organization and dissemination of such assets. Various tools and technology, including open source software has been discovered by libraries to overcome the above situation and these solutions has been accepted as most reliable and cost effective to manage the digital resources word wide.

2. About NIT Rourkela

The National Institute of Technology, Rourkela was founded as the Regional Engineering College, Rourkela on 15th August, 1961. The Institute was declared as the National Institute of Technology with Deemed to be University status on 26th June, 2002 by the Ministry of Human Resource Development, Govt. of India. It is an Institute of National Importance imparting Technical Education both at Under Graduate and Post Graduate level. The NIT Act came into force w.e.f. 15th Aug, 2007. Approximately 6000 and 1000 students and staff serving the institute and it is now ranked almost within 15 among various engineering colleges of India in the survey done by various Indian agencies.

3. About BP Central Library

Biju Patnaik Central Library [BPCL) of NIT Rourkela has been functioning since 1965. Since its inception, it has been rendered essential support to the Institute's teaching, research and development programs. The library presently holds more than 75000 books and 20,000 back volumes of periodicals and equipped with all modern IT based equipments in its four storied building. More than 6000 online journals and databases are accessible which is subscribed by NIT Rourkela and some are provided by INDEST-AICTE consortium as well as BPCL also subscribed 75 Indian Print Journals. Library is having approximately 05 crore rupees budget for books and online journals and

databases. Apart from books and journals, library has also many IS codes, educational video courses, cassettes, CD-ROMS etc.

The BPCL is fully automated with an integrated library software package called Libsys – LSmart and modernized with latest Radio Frequency Identification [RFID) based automation system that facilitates self check-in, self check-out and automatic security system. This technology offers an easy, fast, and an efficient means of tracking, locating and managing of library resources. The RFID system counts more than 1.2 lakhs transactions per year, including issue, return and renewal of books and documents. The working hours of Central Library are from 9.00 hours to 22.00 hours on all working days and on holidays and Sundays it remains open from 9.00 hours to 18.00 hours.

3.1 Major IT based services @ BPCL

The BPCL has, in the recent past, undertaken several innovative initiatives such as Institutional Repository

(IR), Electronic Thesis and Dissertations (ETD), Web-based video courses, Blogs etc.

Library website < <http://library.nitrkl.ac.in/>> is developed and launched with the Joomla, CMS which is the first website of its kind among all NIT'S for regular updates of the library and about resources available to us.

e-book catalogue < <http://opac.nitrkl.ac.in/>> e-books, subscribed by NIT Rourkela have been catalogued which can be searched and directly linked to full text website. It has been developed with most popular library management software Koha.

Institutional Repository (IR) <<http://dspace.nitrkl.ac.in/dspace>> is an open access repository of institutional research output of the faculty, researchers and students. At present, the repository holds nearly 2026 documents consisting of journal papers, conference papers, book chapters and doctoral theses and it is openly accessible and globally.

Electronic Thesis and Dissertations (ETD) <<http://ethesis.nitrkl.ac.in/>> is an institutional open access thesis and dissertations repository of this Institute. This repository is an attempt to make all locally produced theses more visible to global users. It is mandatory for students of all disciplines (BTech, M.Tech. MTech (Res), M.Sc, MA,

MBA and Ph.D) to upload the final version of their theses submitted to the institute as a part of their academic requirements, till date repository holds 3090 ETD.

Web-based Video Courses<<http://dlib.nitrkl.ac.in/videocourse/>> called 'OpenCourseWare' of the national NPTEL programme are available to the students and faculty of NIT Rourkela over the institutional network, The NPTEL video coeres has been upgraded from Phase-I to Phase- II.

Blog <<http://192.168.1.126/blog/>>called '**Library Services Blog**' is an intellectual tool for documentation of information among local users. It alerts and informs patrons about the latest developments in the field of science, engineering, technology, management and humanities.

Open Access to Oriya Books (OAOB)<<http://oaob.nitrkl.ac.in/>> is a non- profitable project, initiated by NIT, Rourkela, Srujanika, Bhubaneswar and PragatiUtkal Sangh, Rourkela. The mission of the project is to preserve the cultural history of Orissa by digitizing old, rare and new literary and other texts written in the Oriya language.

4. DSpace at BPCL

DSpace is an open source repository software package typically used for creating open access repositories for scholarly and/or published digital content. While DSpace shares some feature overlap with content management systems and document management systems, the DSpace repository software serves a specific need as a digital archive system, focused on the long-term storage, access and preservation of digital content. The first public version of DSpace was released in November 2002, as a joint effort between developers from MIT and HP Labs. Following the first user group meeting on March 2004, a group of interested institutions formed the DSpace Federation, which determined the governance of future software development by adopting the Apache Foundation's community development model as well establishing the DSpace Committer Group. In July 2007 as the DSpace user community grew larger, HP and MIT jointly formed the DSpace Foundation, a not-for-profit organization that provided leadership and support. In May 2009 collaboration on related projects and growing synergies between the DSpace Foundation and the Fedora Commons organization led to the joining of the two organizations to pursue their common mission in a not-for-profit called DuraSpace.

Currently the DSpace software and user community receive leadership and guidance from DuraSpace.

4.1 Purpose of DSpace @ BPCL

DSpace at BPCL was installed in 2006 and the main purpose was to archive the following intellectual output of the institute.

- a) **Conference publication:** The main objective is of this DSpace community is to collect all the articles published in the event volumes. It has been made mandatory for all the faculty/staff of the institute to submit the full text paper in this repository if they had attended the event on institute-expenses; hence the total number of this community is healthy in comparison to other.
- b) **Journal Articles:** There are approximately 916 articles archived in the DSpace repository till date which is very less in comparison to above publications which reflects lack of policy on submission of such type of resources.
- c) **Book Chapter:** Total 11 book chapter appears in the collection till date and issue is same as in Journal collection.

4.2 Up-gradation of DSpace @ BPCL

One of the major challenges of open source software is its frequent updates and upgrading the old system is another big challenge, especially if it is managed by in-house technical man powers. The DSpace installed at BPCL was a very old version 1.2.1 on Red Hat version 3.0. Several attempts have been made to upgrade the DSpace with in-house manpower as the hardware and software was in very critical stage due to its life span and other technical issues but nothing worked effectively. However the author and some technical person himself took the initiative and ability to successfully up-grade the one of the oldest DSpace open source repository in India. The entire process of DSpace upgrade is provided below which is very useful for those who wish to upgrade their repository.

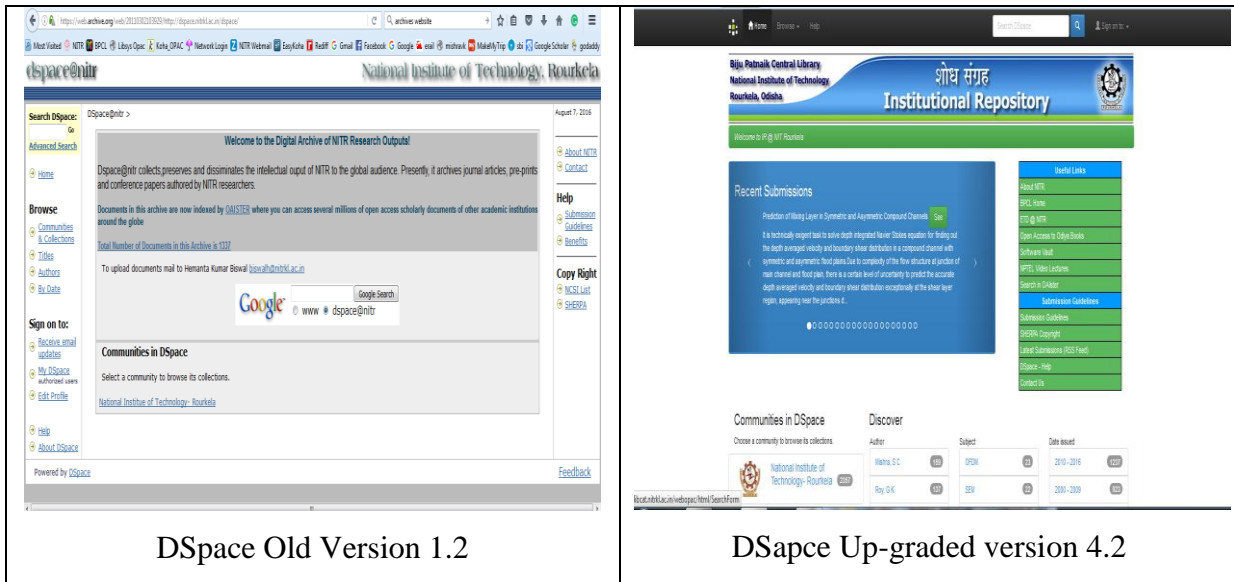


Fig. 1 DSpace look and feel after up-gradation

4.2.1 DSpace Up-gradation from old version to new version Tips & Tricks

Latest version of Dspace 4.2 was installed as per installation instructions and then old database was upgraded (primarily the database) as per the example given below. This is a much easier multi-version upgrade process; however it does not automatically upgrade all of your configurations or customizations so if you have customized your DSpace files and folders this process may not satisfy your requirements.

For example, to upgrade to DSpace 4.x follow the instructions at Installing DSpace. At this point, you'll have a fresh copy of DSpace, but no data. Make a backup copy of your old DSpace database to use this to upgrade your data to the latest version of DSpace, the entire process/step is described in table 1.

Step	Description	Command
1	Use PostgreSQL the "pg_dump" command to backup all your old data	<code>pg_dump -U [database-user] -f my-dspace-db-backup.sql [database-name]</code>
2	Create an exact replica of your old DSpace database (this is mostly just for safety purposes – you don't want to lose any of your Production data)	<code>createdb -U [database-user] -E UNICODE dspace-upgrade (dspace-upgrade is a replica database name)</code>
3	Load the old data into this replica database	<code>psql -U [database-user] -f my-dspace-db-backup.sql dspace-</code>

		upgrade
4	Upgrade this replica database to be compatible with the version of DSpace you wish to upgrade to. This is done utilizing the "database_schema-*.sql" upgrade scripts provided in the [dspace-source]/dspace/etc/postgres/ folder. For example, suppose you are upgrading from DSpace 1.2.x to 4.x. That means you'll be upgrading your database from being 1.2.x compatible to being 4.x compatible. In order to do that, you will run all these database upgrade scripts in order (again this example is just for PostgreSQL)	Upgrade 'dspace-upgrade' database from 1.2.x to 1.3.x compatibility psql -U [database-user] -f [dspace-source]/dspace/etc/postgres/database_schema_12-13.sql dspace-upgrade Similarly you have to run the above command up to latest version which you wish to upgrade.
5	Now you have a 4.x compatible database (in "dspace-upgrade") So, you can dump its data out to a file: For example, to dump "dspace-upgrade" to a file named "my-dspace-upgraded-db.sql"	pg_dump -U [database-user] -f my-dspace-upgraded-db.sql dspace-upgrade (log into as dspace user and save the dump in /home/dspace folder)
6	Delete the default (MAKE SURE IT IS EMPTY) "fresh install" Database on the NEW DSpace	dropdb -U [database-user] [database-name] (log into as dspace user (sudo su dspace) to delete the database)
7	Recreate an empty DB	createdb -U [database-user] -E UNICODE [database-name] (log into as dspace user (sudo su dspace) to create database)
8	Load in the newly upgrade data (from "my-dspace-upgraded-db.sql")	psql -U [database-user] -f my-dspace-upgraded-db.sql [database-name] (log into as dspace user (sudo su dspace) to restore the database and take backup)
9	Next, copy over your old DSpace "Assetstore" directory into the newly installed version of DSpace. This assetstore directory contains all the actual content files	cp -R [path-to-old-dspace]/assetstore/* [path-to-new-dspace]/assetstore/
10	Finally, reindex all of your DSpace content in the newly installed version of DSpace	[dspace]/bin/dspace index-discovery -f
Now fresh installation of a newer version of DSpace with your content migrated. As a final step, you may wish to re-configure or re-customize the fresh installation (based on your old settings in files and folders). Also be sure to review all the Upgrade Instructions for the versions you "skipped over". Sometimes there are important notes or details available for further actions.		

Table 1: Steps of DSpace up-gradation

4.3 Usages Statistics and ranking of DSpace repository @ BPCL

Before up-gradation of DSpace 1.2 there was no mechanism for usage statistics on DSpace items, but the new feature facilitates nice provision for monitoring its usage statistics not only on entire website but at each item level. It provides facility to view the usage statistics for DSpace community, collection and items, one can view the total visits, total visits per month, top country views, top city views and total number of times file downloaded. To calculate the entire usage statistics author has placed free online usages statistics tool which provides good features for analyzing the maximum browsers used, operating system used, location-views, referring sites, search engines etc. figure 2 indicates the type of data which can be retrieved from such free online tools to monitor the usage statistics of the website.

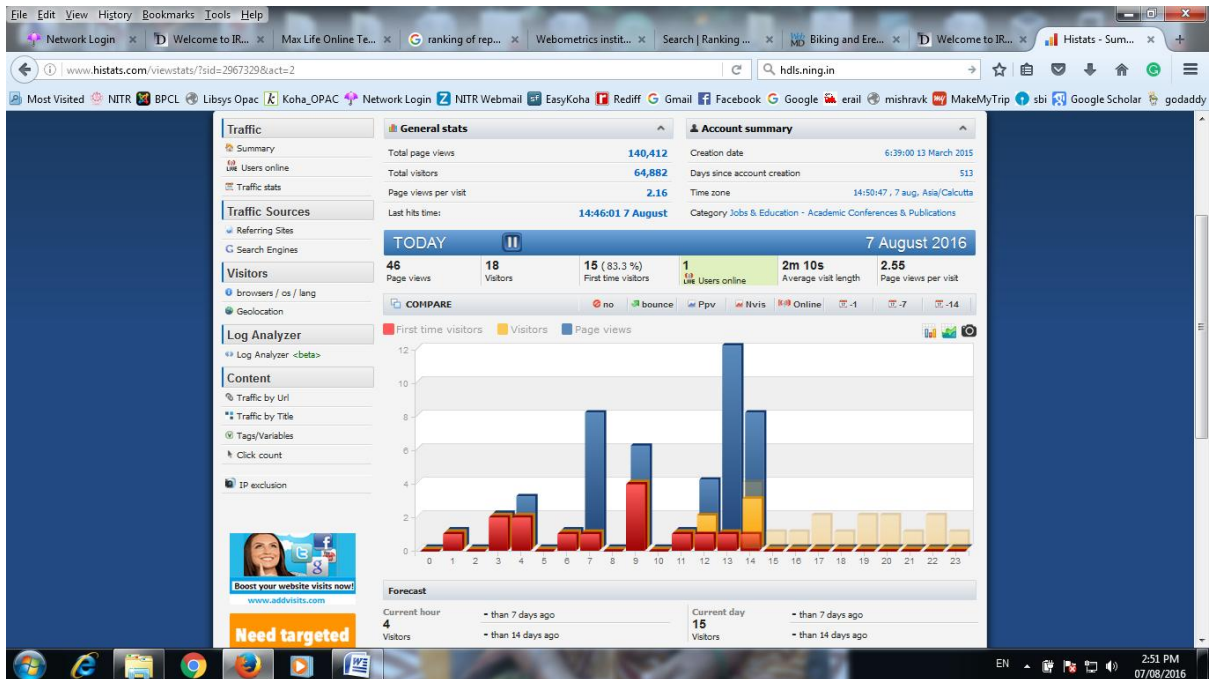


Fig.2 Free <http://www.histats.com> online tool for usages statistics

There are several institutions which monitor and publish the ranking of online repositories out of which Ranking web of repositories site is most popular according to which NIT Rourkela DSpace is placed on 695th Rank in word and 07th in the India which indicates its popularity and usages are very high globally.

5. EPrint @ BPCL

E-print is first and one of the most widely used open digital library software. It is a Web and command-line application based on the LAMP architecture is primarily used for institutional repositories and scientific journals. It has been developed at the University of Southampton, School of Electronics and Computer Science in 2000 and released under a GPL license for building open access repositories that are compliant with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). It offers a significant number of the elements regularly found in record administration frameworks, however is principally utilized for institutional storehouses and experimental scientific journals.

5.1 EPrint @ BPCL

Being a active Institution for OSS movements in India, BPCL started using EPrint to attempt to make all locally produced theses/dissertation/Btech projects more visible to global users in the year 2009. It is made mandatory for students of all courses (BTech, MTech, MTech by Research, MSc and PhD) and from all disciplines to upload the final version of the theses submitted as partial fulfillment of degree. This repository is named as E-thesis@BPCL. Till date the statistical data representing all the thesis are represented in the table 2.

Sl. No	Type of Thesis	No of thesis
1	B-Tech	2469
2	M.A.	60
3	M.Sc.	521
4	M-Tech	1790
5	M-Tech by Research,	197
6	PhD	299

Table 2 representing the thesis type at E-Thesis@ BPCL)

Besides E-thesis@BPCL, library uses e-print for **Open Access to Oriya Books [OAOB]** <<http://oaob.nitrkl.ac.in/>> which was initiated by NIT, Rourkela, Srujanika, Bhubaneswar and PragatiUtkal Sangh, Rourkela in the year 2009. In this repository BPCL preserves the Literary heritage of Orissa by digitizing old and rare books written in the Oriya language. Till date BPCL digitized more than 522 books which is accessible

globally under open access policy of the institute. For digitization of rare and old books library usages high end scanner Bookeye3 and 300 dpi grayscale images quality. The tagged image file format (TIFF) is used for scanning the pages of books and later it is converted into the pdf files for uploading on the e-print software.

6. Koha @ BPCL

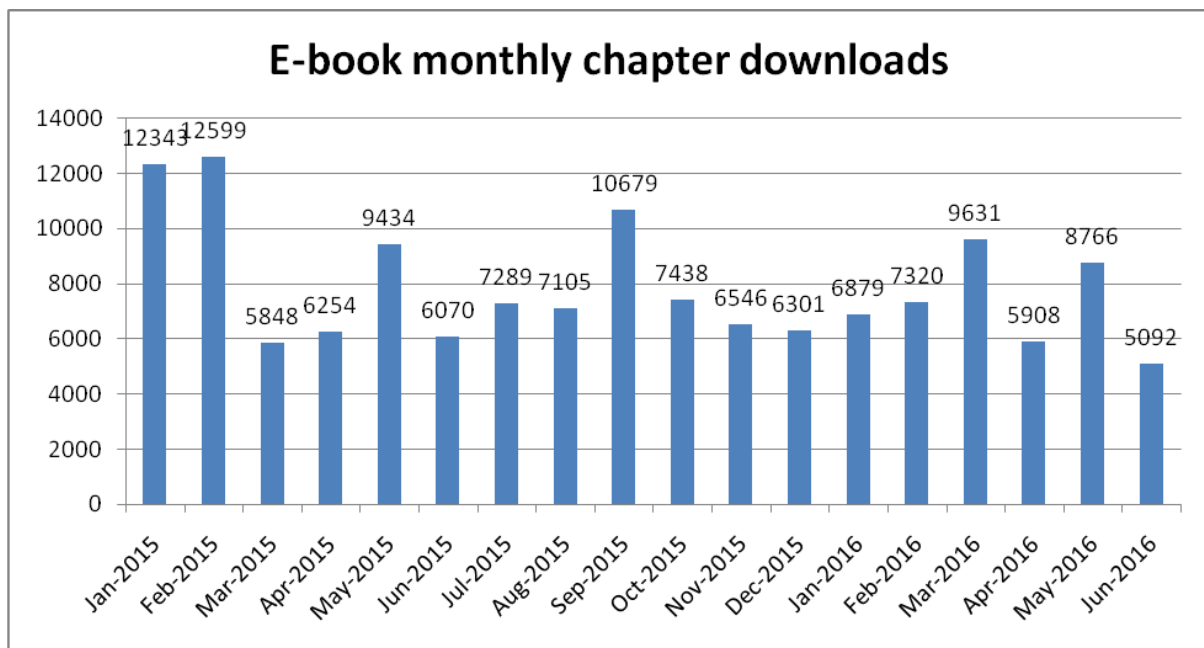
There is continued discussion ongoing at BPCL for full implementation of Koha ILMs but it is not yet done till date mainly because of compatibility related issues with existing RFID devices and its high cost of integration with Koha. However Koha was already implemented at BPCL initially for e-book and other materials like videocassette, CD/DVD etc. which is not catalogued in the existing software Libsys. Till date following item types are available in the Koha at BPCL.

- Videocassette directly catalogued in the Koha
- Bound volume of periodicals migrated from Libsys
- Lecture noted in computer science through MARC provided by Springer
- E-book of Springer in engineering through MARC provided by Springer
- E-books by Elsevier through MARC provided by Elsevier
- CD/DVD catalogue

6.1 Digital resources, bibliographic data conversions at BPCL & e-book statistics

All the above data was refined and converted into MARC before importing into the Koha except videocassette. Bound volumes of journals were migrated from Libsys, the subscribed e-books available at the Springer website in MARC format has been downloaded and little modifications were done as per requirement, some new MARC tag added into the downloaded record is 041\$a and 942\$c, the 041 is to ensure language wise searching of document from Koha web OPAC and 942 is to assign the records as item type (EB=e-book). Similarly, some modifications were done for Elsevier e-books. The process of importing the e-book record is continued at some regular interval of time as BPCL is subscribing the resource continuously.

The digital resources available with BPCL are now visible globally and users of the entire NIT Rourkela community found very useful and helpful in direct access of the e-resources and locating other digital resources like CD/DVD and Videocassettes in the library. The management of these resources with the help of such tools increased almost double of its uses as per the study/statement available on Springer e-books website. The uses statistics of e-books available at BPCL are very huge as indicated in the below chart which indicates implementation of such tools is must in the libraries and information centre.



E-books statistics at BPCL

7. Joomla @ BPCL

Joomla is a free and open-source content management system (CMS) for publishing web content. It is built on a model–view–controller web application framework that can be used independently of the CMS. Joomla is written in PHP, uses object-oriented programming (OOP) techniques (since version 1.5) and software design patterns, stores data in a MySQL or PostgreSQL. Joomla! is an award-winning content management system (CMS), which enables you to build Web sites and powerful online applications. Many aspects, including its ease-of-use and extensibility, have made Joomla the most

popular Web site software available. Best of all, Joomla is an open source solution that is freely available to everyone.

Joomla is used all over the world to power Web sites of all shapes and sizes like Corporate Web sites or portals, Corporate intranets and extranets, Online magazines, newspapers, and publications, E-commerce and online reservations, Government applications, Small business Websites etc. BPCL is maintaining its entire content on Joomla which is windows for the library through which globally people can make themselves aware about BPCL resources. Several full text user manual and guides are managed through Joomla which helps the user to know more about its digital assets and how to access it. A Library used to publish about the latest arrivals and happenings through various digital formats like pdf, MS word, links to videos etc.

8. Greenstone @ BPCL

Greenstone is a suite of software for building and distributing digital library collections. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Read the Greenstone Factsheet for more information.

NIT Rourkela has the subscriptions and collections of many purchased/propriety as well as open source software's which is managed and stored in Greenstone digital library software for dissemination among the entire NIT Rourkela community on the intranet. Apart from software's there are old question papers of undergraduate and postgraduate exams which is also managed and disseminated through this software.

9. Technical and Administrative difficulty in the management of digital assets @ BPCL

Many of the above mentioned software was implemented by independent capacity of Mr. Madhan, ex. Assistant librarian of BPCL later Koha and Joomla were implemented by in-house library professionals. The server is managed by computer center of the institute and

the rest of the technical issues resolved by library professionals as and when require. The major challenge till date library faced is up-gradation of all software due to limitations of the technical skills among library professionals and work culture where a single staff has to perform multiple jobs of varying skills. As far as hardware is concern computer center of the institute managed the same, but it again creates a lot of difficulty in administration and management of the several softwares which lies in the several servers. Due to security reasons library has very limited access to servers, hence they have to approach for any modifications on softwares to computer center's staffs.

The motivations and some administrative policy are also lacking because of which the item count of a very old repository is less. The use of so many software like e-print, Koha, Joomla, DSpace, libsys etc. Again creating lot of confusion among users and its technical management. Library in-house manpower is regularly putting their best efforts for improvements and up-gradations of the system. Library staff is responsible for submission of digital items like e-theses, dissertation, projects, Articles, conference proceedings etc. Contradictory to submission by users hence there is lacking in proper policy and workflow in digital document collections, preservation and dissemination exist at present.

10. Conclusion

Above paper reflects that there are lots of open source solutions (OSS) available for management of digital resources in the library. Digital assets available in any media like files, audio, video, etc can be managed, preserved and accessed globally with the help these freely available globally recognized OSS. Proper administrative policy and technical manpower are very much needed in the library for overall development of digital resources and handling the technical difficulties in running the software's. Using many such solutions may be not good for its user and single point access of all such resources is badly needed under such type of environment.

In house training of library staff on such tools and technology is very essential to cater the future requirement for up-gradation or switching from one software solution to another. Management of such resources by in-house human resources saves a huge amount of the institute and build very good professional image among the entire

community. Hence authors would like to recommend the adoption of such solutions especially for the developing country-libraries with the help of community support so that cost of implementation can be reduced in overall digital asset management.

11. References

- About Greenstone, <http://www.greenstone.org/> (accessed on 09/08/2016).
- About Joomla, <https://www.joomla.org/about-joomla.html> (accessed on 28/05/2016).
- Annual Report, National Institute of Technology Rourkela, 2014-15.
- Biju Patnaik Central Library Website: Joomla, <http://library.nitrkl.ac.in/> (accessed 2016).
- DSpace, <http://www.dspace.org/> (accessed 15/05/2016).
- DSpace, <https://en.wikipedia.org/wiki/DSpace> (accessed 2016)
- Ebook and additional resources catalogue: Koha, <http://opac.nitrkl.ac.in/> (accessed on 06/08/2016)
- EPrint, <http://www.eprints.org/uk/> (accessed on 04/05/2016)
- EThesis @ NIT Rourkela: eprint, <http://ethesis.nitrkl.ac.in/> (accessed 08/08/2016).
- Institutional repository: DSpace, <http://dspace.nitrkl.ac.in/dspace/> (accessed on 07/07/2016).
- Koha library software: <https://koha-community.org/> (accessed on 23/05/2016).
- National Institute of Technology Rourkela, <http://nitrkl.ac.in/> (accessed 2016).
- Open access to odia books: eprint, <http://oaob.nitrkl.ac.in/> (accessed 07/08/2016)
- Software vault and library resources: greenstone, <http://dlib.nitrkl.ac.in/> (accessed on 05/07/2016)