

Institutional Repository at Central Library IIT Kharagpur: An Overview

¹Debasis Das, ²Parnab Chatterjee

¹ Junior Project Officer, National Digital Library Project, IIT Kharagpur

² School Librarian, Janardandih High school, West Bengal

Abstract - An institutional repository is an online resource for the storing in digital form of academic materials, such as thesis, dissertations and research articles, on behalf of a university or other institution. This paper focuses on the requirements, functions and use of digital preservation in an institutional repository context. The paper highlights on various feature and importance of both Institutional Digital Repository and Open Source Digital Library Software that is DSpace. The Central Library IIT, Kharagpur uses open sources digital library software namely DSpace. The IDR of Indian Institute of Technology Kharagpur (IITKGP) collects preserves and makes available in digital format of the scholarly output of IIT Kharagpur community. Its interface provides for an easy self-archiving by faculty, and organizes the documents in logical, easily retrieved fashion.

Key Words - Institutional Digital Repository; Digital Library Software; DSpace; Digital Preservation; Bookeye 4

I. INTRODUCTION

An Institutional repository is a mechanism for managing and storing digital content. Repositories can be subject or institutional in their focus. Putting content into an institutional repository enables staff and institutions to manage and preserve it, and therefore derive maximum value from it. A repository can support research, learning, and administrative processes. Repositories use open standards to ensure that the content they contain is accessible in that it can be searched and retrieved for later use. The use of these agreed international standards allows mechanisms to be set up which import, export, identify, store and retrieve the digital content within the repository.

Digital repositories may include a wide range of content for a variety of purposes and users. What goes into a repository is currently less an issue of technological or software ability, and more a policy decision made by each institution or administrator. Typically content can include research outputs such as journal articles or research data, e-theses, e-learning objects and teaching materials, and administrative data. Some repositories only take in particular items (such as theses or journal papers), whilst others seek to gather any credible scholarly work produced by the institution; limited only by each author's retained rights from publishers. However, some more complex objects (websites, advanced learning objects, 3D topographical representations and other data sets) do present a technological challenge.

The origin of the notion of an "institutional repository" are twofold: Institutional repositories are partly linked to the notion of digital interoperability, which is in turn linked to the Open Archives Initiative (OAI) and its Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The OAI in turn had its roots in the notion of a "Universal Preprint Service" since superseded by the open access movement.

Institutional repositories are partly linked to the notion of a digital library - i.e. collecting, housing, classifying, cataloguing, preserving, and providing access to digital content, analogous with the library's conventional function of collecting, housing classifying, preserving and providing access to analog content.

Availability of Open Source System Software's has facilitated the setting up of repositories. The commonly used open source software's for creation of the Institutional Digital Repository are DSpace and Eprints, others being MyCore, Fedora, Greenstone, Archimed, ARNO, CDSware, Tor, and OPUS.

II. BENEFITS OF AN IR

The benefits of repositories to institutions and individuals are numerous and can be grouped into the following categories:-

- Digital preservation of the documents.
- Centralization and storage of all types of institutional output, including unpublished literature.
- Support for learning and teaching. Links may be made with the virtual teaching environment and library catalogues.
- Standardization of institutional records. The compilation of an 'Institutional CV' and individual online dossiers linked to the full text of articles becomes possible.
- Ability to keep track of and analyze research performance.
- Breaking down of publishers' costs and permissions barriers.
- Alleviation of requirement to trust publishers to maintain information in the long term, without any commercial benefit for the authors.
- Promotion of a philosophy of wider communication.
- Online access for the Institutional Research Publication.

- Opportunities for new forms of scholarly communication.

III. CHARACTERISTICS OF INSTITUTIONAL REPOSITORY

Institutional Repository has four characteristics as defined by Johnson (2002). They are mentioned below:

- Institutionally defined (as opposed to discipline- or subject-focused)
- Scholarly (containing the products of faculty, research staff, and students)
- Cumulative and perpetual (the content will be preserved on a long-term basis);
- Open and interoperable (attentive to the Open Archives Initiative—Protocol for Metadata Harvesting).

IV. GROWTH AND DEVELOPMENT OF IRs IN INDIA

Few years ago, Institutional repository initiatives were initiated only in some developed countries. In India, Institutional repository movement was started in 2004 and is gaining momentum and new initiatives are emerging around the country. Institutional repositories are already well established in many IITs, IIMs and other institutes having national importance. Many universities and research institutes in India have developed institutional repository for archiving documents of their own. It is not only limited to Science and Technology but also on other disciplines. More than 60 academic and research institutions have set up their Institutional Repositories as indicated by ROAR (Registry of Open Access Repository) and DOAR (Directory of Open Access Repository) viz., IISc, IIMK, ISI, NCL, NIO, RRU, NAL, NIT and so on. There are a few institutions have not registered in ROAR or DOAR. Recently, MHRD (Ministry of Human Resource Development) has also advised all INDEST members (Indian National Digital Library in Engineering, Sciences and Technology) to set up institutional repository using open source software. University Grants Commission (2005) has already developed a policy document on building University level Institutional Digital Repository (http://www.ugc.ac.in/new_initiatives/etd_hb.pdf) in India. UGC has recommended that all the universities should set up these repository to facilitate e-submission, archiving, maintenance and access to these repository at the university level. University Grants Commission (2005) enacted "UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005" to strengthen national capability of producing electronic theses and dissertations, and, to maintain university-level and national level databases of theses and dissertations. This Regulation proposed two sets of planned actions, such as:

Creation of Indian National Theses Database (INTED)

Submission of PhD Theses in Electronic Form

Another government organization, National Knowledge Commission (2007) strongly advocates open access to public-funded research literature and recently has taken initiative for building nationwide institutional Repository. Bangalore declaration (2006) drafted a model National Open Access Policy for Developing Countries also support this view and advocated for open access institutional repository. Some professional associations and societies like Developing Library Network (DELNET), INFLIBNET (Information Library Network) are also involved in modernization of libraries, training and setting up the IRs.

Repository: Purpose of use

- Providing new opportunities for the archiving and preservation of valuable digital works
- Reduce duplication of records and inconsistencies in multiple instances of the same works
- Offer greater flexibility and integration than other mechanisms of disseminating scholarly works with more coherent security and preservation of digital materials
- Assists research collaboration through facilitating free exchange of scholarly information
- Increase the visibility, reputation and prestige of the institution

Repository: a store house

Generally it has been found that institutional repository contents full-text contents of journal articles, conference papers, book chapters, monographs, research reports, project reports, theses, dissertations, patents, presentations, computer programs, tutorials, convocation addresses, audio materials, video materials, course materials, multimedia materials, handbooks, data books, technical manuals etc.

Published papers (preprints, post prints, conferences)

V. DSPACE

DSpace 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 archives system, focused on the long-term storage, access and preservation of digital content. DSpace is the software of choice for academic, non-profit, and 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.

Reasons to Use DSpace

- Largest community of patrons and developers worldwide.
- Free open source software.
- Completely customizable to fit our needs.
- Used by many educational, government, private and commercial institutions.

- Can manage and preserve all types of digital content.

Each DSpace site is divided into communities; these typically correspond to a laboratory, research center or department. Communities contain collections, which is a grouping of related content. Each collection is composed of items, which are the basic archival elements of the archive. Items are further subdivided into bundles of bitstreams. Bitstreams are, as the name suggests, streams of bits, usually ordinary computer files. Bitstreams that are somehow closely related are organized into bundles, for example HTML files and images that compose a single HTML document. The data model supports multiple inclusion at all levels; that is, an item may belong to more than one collection, and a collection may be in more than one community. Each item has one qualified Dublin Core metadata record. Other metadata might be stored in an item as a serialized bitstream, but we store Dublin Core for every item for interoperability and ease of discovery. The Dublin Core may be entered by end-users as they submit content, or it might be derived from other metadata as part of an ingest process.

VI. INSTITUTIONAL REPOSITORY AT IIT KHARAGPUR

The central library of IIT Kharagpur is the first of the eight IIT Libraries in India to set up an Institutional Repository in which the digital documents are organized in to three major categories of collection, such as Administration, Library and Departments / Centers / Schools under the common heading as "Community & collection". Thus collection against each of the above mentioned Community reflects the core area of activity of the community concerned as its main subject of collection. New communities will be created as and when the same is required to give birth to new and emerging subject areas and activities in this Institute.

Materials in the Repository

Central Library IIT, Kharagpur have included all the departments/ centers and administration as individual community. Registered faculty members, research scholars and research assistants working at IITKGP may submit their documents to the Institutional Repository. There are different types of community class in this repository, for example- PhD These of IIT Kharagpur, D.Sc these of IIT Kharagpur, Download full text. There is also a repository link to access repository all over the world. If you want to browse the collection of a community you have to select the particular community.



Fig 1 Main page of institutional Repository at IIT Kharagpur with Communities

VII. SYSTEM REQUIREMENT

For implementation an institutional repository we need two types of requirement.

- Hardware
- Software

A) Hardware Requirement

Minimal DSpace Production system requirements

2 GB of Random Access Memory (RAM)

1GB for Tomcat (e.g. "TOMCAT_OPTS=-server -Xms1024M -Xmx1024M -XX:MaxPermSize=128M -Dfile.encoding=UTF-8") 1GB for Database (PostgreSQL or Oracle).

20 GB of Storage (or roughly enough storage for all the files you wish to store in DSpace)

Mid-range DSpace Production system

- 4 GB of Random Access Memory (RAM)
- 200 GB of Storage (or roughly enough storage for all the files you wish to store in DSpace)

High End DSpace Production system requirements:

Quad Core processor

- 8GB of Random Access Memory (RAM)
- 73 GB 15,000 rpm network disks in RAID accessible over a gigabit connection for storing the database and indexes.
- 7,400 rpm network disks in RAID accessible over a gigabit connection for storing the data whose size can be easily expanded.



Fig 2 High speed Bookeye 4 Scanner with OCR Software (to digitize the print document)

Table 1 Specification of the Scanner Bookeye 4

Type	BE4-SGS-V2
Method of Scanning	Scanning through mirror rotation
Maximum Scan area	620 X 460 mm
Resolution	150/ 200/ 300/ 400/ 600 dpi
Format	Flat mode, Book fold correction, V mode
Scanning format	Maximum, Auto, Crop and Deskew, DIN
File format	Jpeg, pnm, tiff, pdf
Colour Mode	24 bit, 8 bit, Gray scale, Binary, Photo

B) Software requirement

DSpace:

Download from <http://www.dspace.org>

In addition, some other software was needed to run DSpace including:

UNIX-like OS or Microsoft Windows

Oracle Java JDK 7 (standard SDK is fine, you don't need J2EE) or OpenJDK 7

Oracle's Java can be downloaded from the following location:

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>.

Apache Maven 3.x (Java build tool)

Maven is necessary in the first stage of the build process to assemble the installation package for your DSpace instance. It gives you the flexibility to customize DSpace using the existing Maven projects found in the `[dspace-source]/dspace/modules` directory or by adding in your own Maven project to build the installation package for DSpace, and apply any custom interface "overlay" changes. Maven can be downloaded from the following location: <http://maven.apache.org/download.html>

Apache Ant 1.8 or later (Java build tool)

Apache Ant is required for the second stage of the build process. It is used once the installation package has been constructed in `[dspace-source]/dspace/target/dspace-<version>-build` and still uses some of the familiar ant build targets found in the 1.4.x build process. Ant can be downloaded from the following location: <http://ant.apache.org>

Relational Database: (PostgreSQL or Oracle)

PostgreSQL 8.4 to 9.1

PostgreSQL can be downloaded from <http://www.postgresql.org/>. It is highly recommended that you try to work with Postgres 8.4 or greater, however 8.3 should still work. Unicode (specifically UTF-8) support must be enabled. This is enabled by default in 8.0+. Once installed, you need to enable TCP/IP connections (DSpace uses JDBC):

In `postgresql.conf`: uncomment the line starting: `listen_addresses = 'localhost'`. This is the default, in recent PostgreSQL releases, but you should at least check it.

Then tighten up security a bit by editing `pg_hba.conf` and adding this line: `host dspace 127.0.0.1 255.255.255.255 md5`.

This should appear *before* any lines matching all databases, because the first matching rule governs.

Then restart PostgreSQL.

Apache Tomcat 7 or higher

Tomcat can be downloaded from the following location: <http://tomcat.apache.org>.

DSpace Installation Prerequisites

Prerequisites

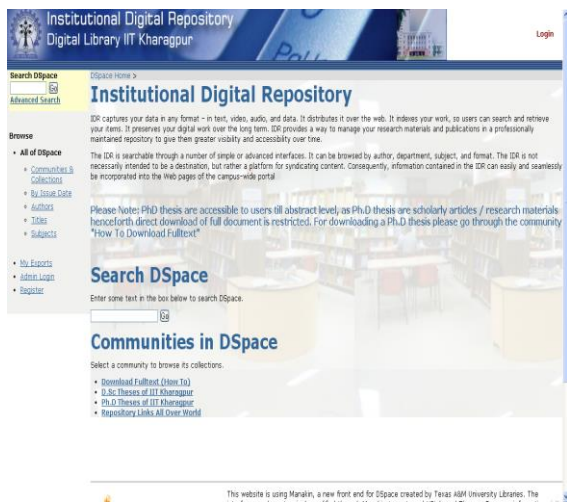
The list below describes the third party components and tools we'll need to run a DSpace server. These are simply recommendations based on our setup at MIT, since DSpace is built on open source, standards-based tools, and there are numerous other possibilities and setups.

1. UNIX-like OS or Microsoft Windows

2. Oracle Java JDK 7 (standard SDK is fine, you don't need J2EE) or OpenJDK 7
3. Apache Maven 3.x (Java build tool)
4. Apache Ant 1.8 or later (Java build tool)
5. Relational Database: (PostgreSQL or Oracle)
6. Servlet Engine (Apache Tomcat 7 or later, Jetty, Caucho Resin or equivalent)
7. Perl (only required for [dSPACE]/bin/dSPACE-info.pl)

VIII. STEPS FOR SUBMITTING CONTENT IN THE IR, IIT KHARAGPUR

1) At first go to the Home Page of IDR



2) Log in to DSpace



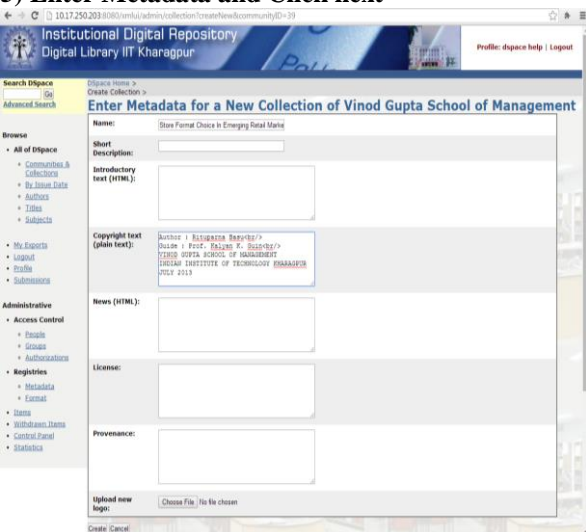
3) Click on Communities and Collection



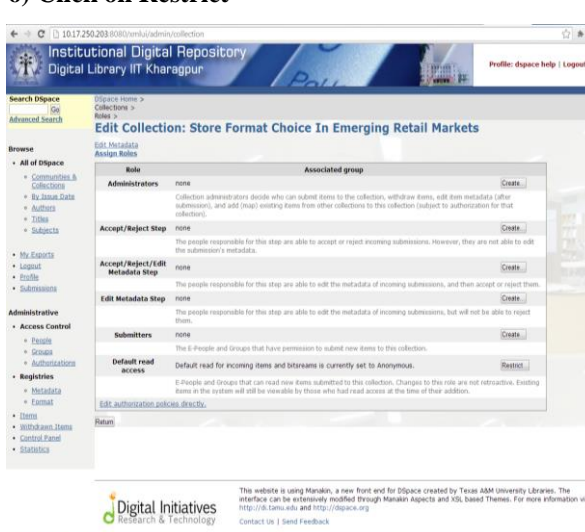
4) Select Sub Community



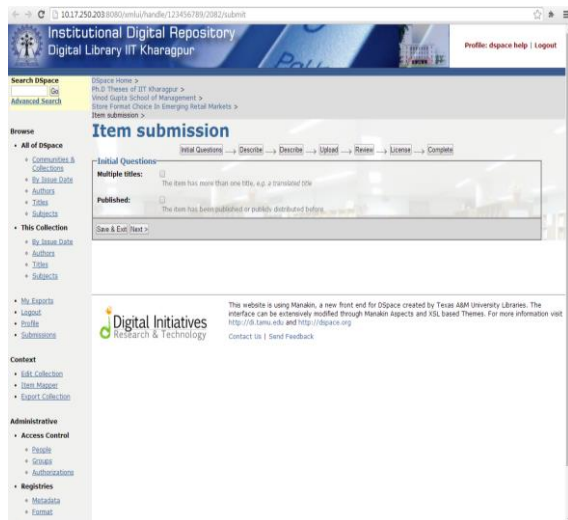
5) Enter Metadata and Click next



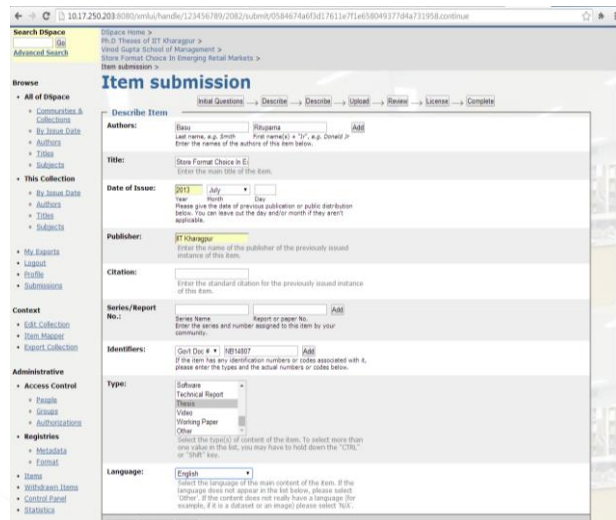
6) Click on Restrict



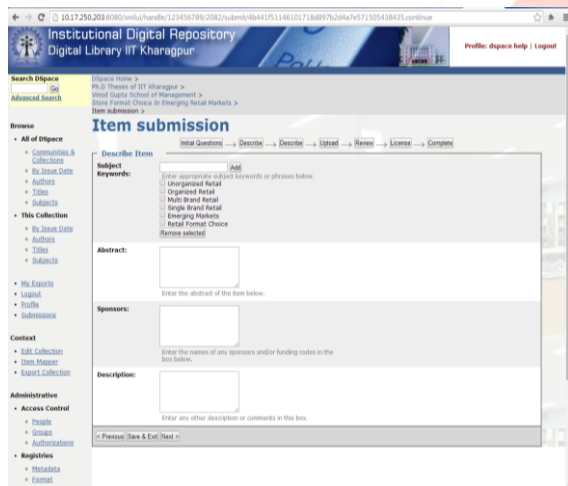
7) Check on published box and click next



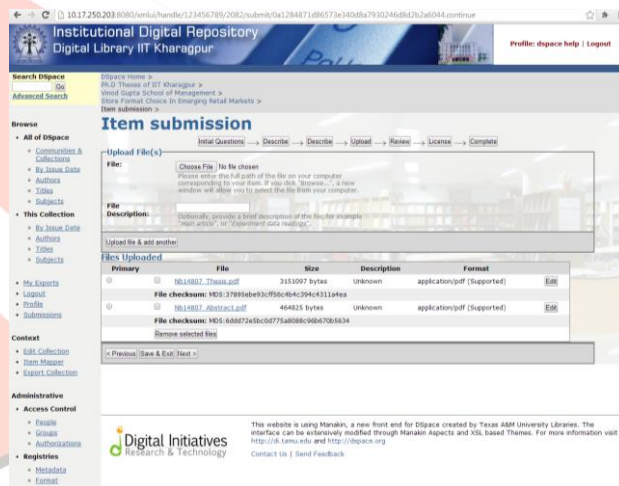
8) Enter the Metadata and click next



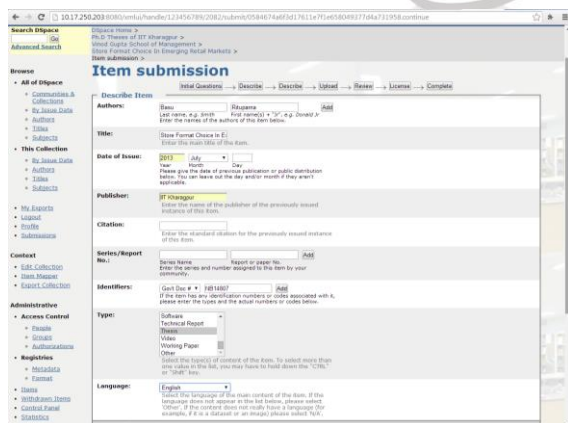
9) Enter the keyword and click next



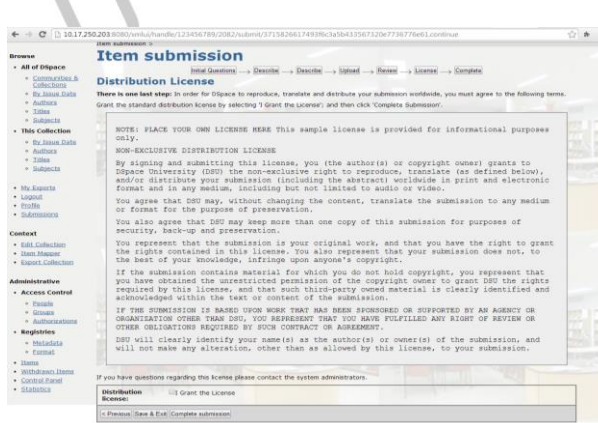
10) Upload files and Click next



11) Check Metadata and click next

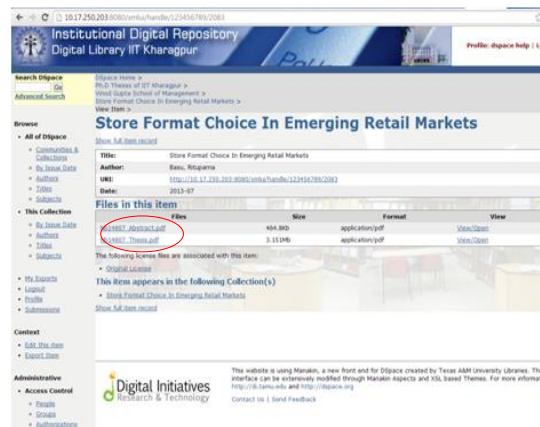
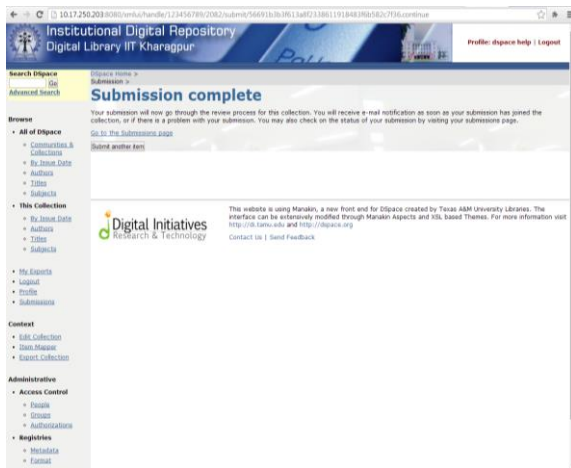


12) Click on grant Institution's license



13) Submission Complete

14) See the Repository in DSpace



IX. BENEFITS OF IDR AT IIT KHARAGPUR

There are many benefits of the IDR at IIT Kharagpur such as

- A repository enables the institute to publish its own scientific research and to make it available to all of its researchers.
- It can be access 24 hours a day.
- It has a safe, backed-up and secure place to store institution scholarly works.
- Make possible easy access to Electronic Theses and Dissertations (ETD). However only print version of theses, dissertations were available a few years back.
- More than one person can access a particular document at the same time.
- Preserves and provides long-term access to the scholars' research output.
- Provide new opportunities for the archiving and preservation of valuable digital works.
- Provides a global platform for local research and hence improved visibility
- Reduce duplication of records and inconsistencies in multiple instances of the same works.
- The quality of the institute's intellectual output can be disseminated effectively and efficiently
- The visibility of the academic output of Indian Institute of Technology will be increased.

X. CONCLUSION

IIT Kharagpur's central library by using Dspace collects and preserves the research output in digital format. For easy retrieval, central library enables the Institute's community to deposit their preprints, post prints and other scholarly publications using a web interface. The access of 'Institutional Digital Repository' is restricted within the IIT Kharagpur campus LAN only. IIT Kharagpur research community only can submit their documents to this repository. It is not just enough to create the IDR. When an institution collects and shares its output, the members of the institution will be benefited and also making the world academically richer by allowing scholarly communication. Many libraries worldwide are involved in interoperable IDRs to improve access to local research for global access.

XI. REFERENCES

- [1] Halling, C. (2014). "History of the Institutional Repository". eHow Contributor , Retrieved from http://www.ehow.com/about_6693254_history-institutional-repository.html (accessed on 19.05.14)
- [2] Li, Y. (2011). "Institutional Repositories and Digital Preservation: Assessing Current Practices at Research Libraries." D-lib Magazine, 17(5). Retrieved from <http://www.dlib.org/dlib/may11/yuanli/05yuanli.html>
- [3] Lynch, C. A. (2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age". Association of Research Libraries. Retrieved from <http://www.arl.org/bm~doc/br226ir.pdf>.
- [4] Roy, B. K., Mukhopadhyay, P. and Biswas, S. K. (2011). "An Analytical Study of Institutional Digital Repositories in India". Library Philosophy and Practice (e-journal) Retrieved from <http://digitalcommons.unl.edu/libphilprac/692/>
- [5] Singh, Vishwanath Pratap. (2013). "Institutional Digital Repository (IDR) at Central Library, Indian Institute of Technology Kharagpur: A Case Study" Retrieved from <http://lsrj.in/UploadedArticles/169.pdf>
- [6] Sutradhar, B. (2010). "Open Access and Institutional Repositories: A Case Study of Indian Institute of Technology", Kharagpur Retrieved from <http://mgcl.iitr.ac.in/seminar18-19nov2010/>
- [7] Sutradhar, B. (2006). Design and development of an institutional repository at the Indian Institution of Technology Kharagpur. Program: electronic library and information systems, 40(3), pp. 244-255. Retrieved from <http://www.emeraldinsight.com/Insight/ViewContentSeveral?Filename=/published/emeraldfulltextarticle/pdf/2800400305.pdf> (accessed on 19.05.14)
- [8] Wheatley P. (2004). "Institutional Repositories in the context of Digital Preservation". Digital Preservation Coalition. Retrieved from http://www.dpconline.org/index.php?option=com_docman&task=doc_download&gid=34

Web Bibliography

- [1] Dspace .<http://dspace.org/> (accessed on 20.05.14)
- [2] Dspace at IITkgp Library.<http://library.iitkgp.ernet.in/> (accessed on 20.05.14)
- [3] Institutional Repository. http://en.wikipedia.org/wiki/Institutional_repository (accessed on 23.05.14)
- [4] Open source Software. http://en.wikipedia.org/wiki/Open_source_software (accessed on 23.05.14)
- [5] What is repository? <http://www.rsp.ac.uk/start/before-you-start/what-is-a-repository/>(accessed on 20.05.2014)

