Data Storage And Retrieval

B.E. Dissertation

Submitted in partial fulfillment of the requirement of

University of Mumbai

For the Degree of

Bachelor of Engineering

(Electronics & Telecommunication Engineering)

by

Faiz Ahmed Khatri(14ET27)

Mazharuddin Shaikh (11ET51)

Under the guidance of

Asst. Prof. Rahul B.Khadse

Department of Electronics and Telecommunication Engineering

Anjuman-I-Islam's Kalsekar Technical Campus,

Sector 16, New Panvel, Navi Mumbai - 410206

(Affiliated to University of Mumbai)

Academic Year: 2018-19

1

Certificate

This is to certify that, the dissertation titled

"Data Storage and Retrieval"

is a bonafide work done by

Faiz Ahmed Khatri(14ET27)

Mazharuddin Shaikh(11ET51)

and is submitted in the partial fulfillment of the requirement for the

degree of

Bachelor of Engineering

in

Electronics & Telecommunication Engineering

to the

University of Mumbai

_____   _____

Supervisor   **Examiner**

_____   _____

Head of Department   **Director**

Project Report Approval For Bachelor Of Engineering

This is to certify that the dissertation entitled **"Data Storage and Retrieval"** is a bonafide work done by **Faiz Ahmed Khatri and Mazharuddin Shaikh** under the guidance of **Prof.Rahul B.Khadse.**This dissertation has been approved for the award of **Bachelor's Degree** in **Electronics & Telecommunication Engineering**, **University of Mumbai**.

Examiners:

_____

Examiner

_____

Supervisor

3

Declaration

We declare that this written submission represents ours ideas in ours own words and where others ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in ours submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Faiz Ahmed Khatri(14ET27)

Mazharuddin Shaikh(11ET51)

Date:April 20, 2019.

Acknowledgement

We are highly grateful to the **PROF. AFZAL SHAIKH, HOD of Electronics & Telecommunication Department, Kalsekar Technical Campus (New Panvel)**, for providing this opportunity to carry out the project. We would like to express our gratitude to other faculty members of Electronics & Telecommunication Engineering Department for providing academic inputs, guidance and encouragement throughout this period. We would like to express our deep sense of gratitude and thank to **ASST. PROF. Rahul B.Khadse**, for the wise council and able guidance it would have not been possible to carry out this project in this manner. Finally, we express our indebtedness to all who have directly or indirectly contributed to the successful completion of this project.

Faiz Ahmed Khatri(14ET27) _____

Mazharuddin Shaikh(11ET51) _____
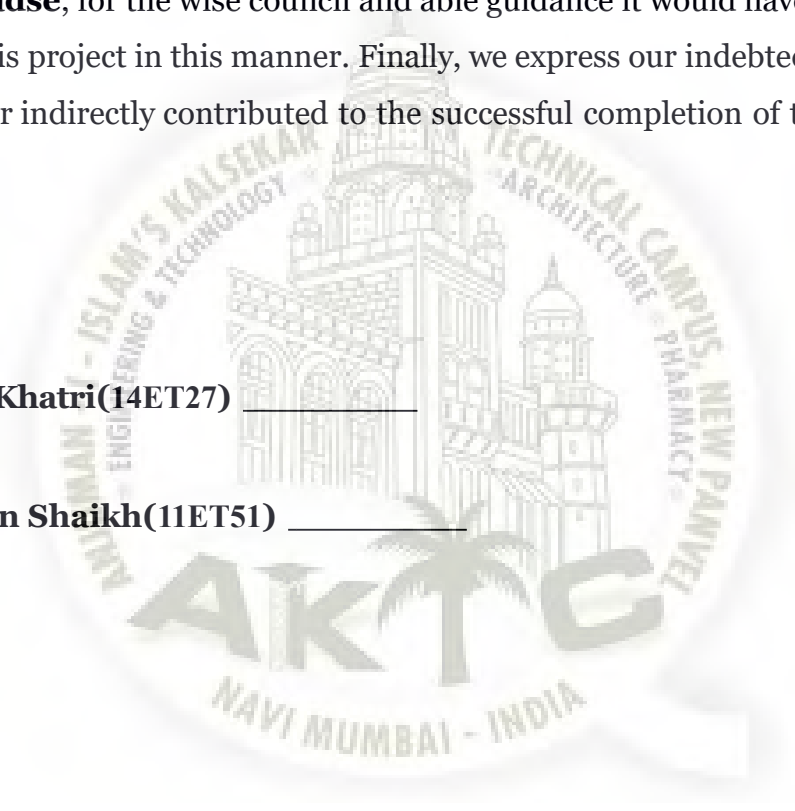
Abstract

Popular use of various Data repository and Web Development are getting more significant than ever. The purpose of this project is to Upload and Download the Documents. Foe this Project we are using HTML, CSS, and Bootstrap for Frontend And for backend we have PHP and MySQL for Database connectivity.

TABLE OF CONTENTS

Chapter
INTRODUCTION

Due to rapid-changing technologies the user expectation has been altering for entertaining the adopted services to these technologies. User now desire and expect content in transportable form, which can be utilized in various digital environment. In a digital age everyone wants the information service providers and information centers for providing the services in such a way that the end users may manipulate the content as they desire. Therefore, in the present scenario, institutional repositories are become an indispensable component for information and knowledge sharing in the scholarly world.

Concept of an Institutional Repository

Institutional Repositories are proliferating, as they become an indispensable component for information and knowledge sharing in the scholarly world. With the increasing Institutional Repositories worldwide, a new phase is emerging with the time solely as a place to store, organize and access content. An institutional repository is not simply a fixed set of software and hardware but an essential tool for intellectual life and scholarship of the institute where its contents can be represented, documented and shared in digital form. The institutional repository is a digital substitute for traditional bulk scholarly publication venues.

9

The institutional digital repository provides a method for capturing and maintaining today's electronic detritus so that tomorrow's scholars can understand the thinking behind the published record. In addition, the institutional repository provides a way for an institution to capture the more polished electronic works i.e. books, article, dissertation, technical papers/report etc. and to graduate access by the organization, researchers worldwide.

The development of institutional repositories emerges as a new strategy that allows institution to apply serious, systematic, leverage to accelerate changes taking place in scholarship and scholarly communication. Many technology trends and development efforts together are making this strategy possible. Technology costs, especially storage and networking costs, have dropped significantly so that repositories are now affordable, therefore, repositories are now affordable where standards like the OAI-MHP (Open Archives Metadata harvesting protocol) are now in place. The convergence of technology developments and other initiatives in recent years has made institutional repositories possible. Awareness of the needs and challenges of digital preservation has accelerated. Development in web publishing such as open archives initiatives, open access journals and disciplinary archives are suggesting new opportunities to enhance scholarly publishing.

10

What is Repository?

An institutional repository is a web-based database (repository) of scholarly material. It is institutionally defined as depend upon the each institution. It could be cumulative and perpetual (a collection of record). It must be open and interoperable (using OAI-complaint software). The institutional repositories are collecting, storing and disseminating digital resources and long term preserving these digital materials.

An institutional repository is a set of services that an institute/university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.

An approved institutional repository provides service having the collaboration among librarian, information technologists, archives and record managers, administrators and policy makers.

Definition: Institutional Repository

As per the Oxford English Dictionary repository is a −vessel receptacle, chamber in which things are or may be placed, deposited or stored‖.

As per Wikipedia −A repository is a central place where data is stored and mined. A repository can be a where multiple databases or files are located for distribution over a network or a repository can be a location that is directly accessible to the user without having to travel across a network‖

11

Benefits of an Institutional Repository

Institutional Repositories by capturing, preserving and disseminating collective intellectual capital, serves as meaningful indicators of an institution's academic quality. It has been seen that much of intellectual output and value of an institution's intellectual property is defused through thousands of scholarly journals emerging from an institutional repository.

An Institutional Repository concentrates the institutional product credited by academic or other institutions researchers, making it easier to demonstrate its scientific, social and financial values.

The Institutional Repository's increased visibility reflects at a high quality of scholarship. Its demonstration of value can be translated into tangible benefits including the funding from public and private sources that drives in part from an institution status and reputation.

Key Benefits of Institutional Repositories

Provide long term solution for scholarly life;

Improve scholarly communication;

Remedying the weakness of self archiving i.e. lack of proper security, long term preservation, wasting faculty time;

Extending the work of disciplinary repositories; Improve online teaching both on campus and via net. Support broad pan-institutional efforts Offers direct and immediate benefits to each subinstitution that cordially implemented the repository.

Basic Elements of Institutional Repository

Institutionally Defined

Institutional Repositories represent the historical and tangible and embodiment of the intellectual life and output of an institution. It captures the original research and other intellectual property in various fields of knowledge that well suit for the institution. It is one of the significant indicators of the institution's academic quality. Therefore, institutional repositories are usually defined institutionally.

13
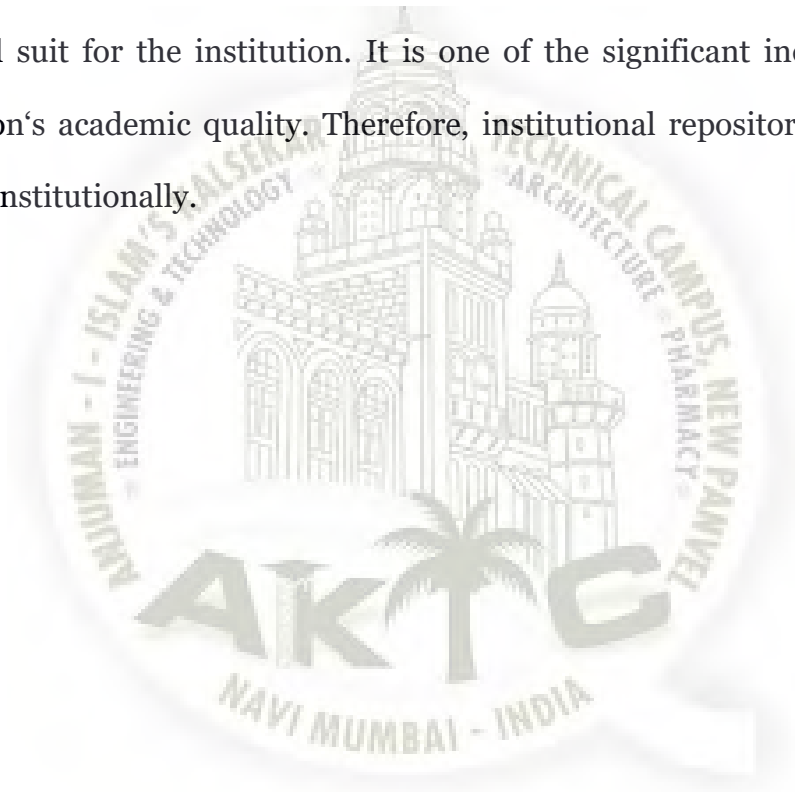
Scholarly Contents

The contents of an Institutional Repository are usually seen as pre prints, pre-reviewed articles, monographs enduring teaching materials, data-sets, conference papers, electronic thesis and dissertation, gray literature and other work in progress. Thus, the contents of an institutional repository would be a collection of scholarly contents.

Cumulative and Perpetual

The collection of a repository is not a one day submission work and not a single bundle of article but contents collected and being uploaded day-to-day would be both cumulative and maintained in perpetuity.

Interoperability and Open Access

To provide access to research community who belongs to outside institution the institutional repositories systems must be able to support interoperability in order to provide access via multiple search engines and other discovery tools at a time. At the same time as an intuitional repository is defined institutionally, the access to this repository by its candidate must be open and interoperable.

Digital Repository Software:

In History, IBM issued Digital Library Software in 1991 to manage collection of digital files. IBM groundbreaking technology grappled with key issue of storage, maintenance, retrieval and display digital content. This was the first effort towards the digital repository software and it showed path to other.

Digital Repository Software: Open Source

There are number of software's available for creating/developing institutional digitals repositories with readily available open source cord; here some brief notes are given for some Open Repository Software;

DSpace

DSpace (http://www.dspace.org) was developed jointly by the MIT library and Hewlett-Packard (HP) Labs. DSpace modestly describes itself as a groundbreaking digital repository system. It captures, stores, indexes, preserves and redistributes an organizations research material formats. DSpace supports institutional repositories and electronic records management. DSpace is being popularly used worldwide to meet many digital archiving needs.

Eprints

Eprints (http://www.eprints.org) is the original digital repository software developed by the University of Southampton to manage an open archive. Eprints was the Open Archives Initiative (OAI) –Complaint repository software. It typically supports collections of pre-prints and technical reports often subject based in scope.

Fedora

Fedora (Flexible Extensible Digital Object and Repository Architecture) is a digital repository system developed jointly by Cornell University Information Science and University of Virginia Library as project. The Fedora projects goal is to provide open-source repository software and related services to serve as the foundation for many different types of Information Management system. Fedora is not a complete system such as DSpace and Eprints whereas it provides an infrastructure upon which services can be developed. It also promotes the buildings of customs tools to expose the repository in creative ways.

Greenstone:

Greenstone (http://www.greenstone.org) is software for building and distributing digital library collections. This software is produced by the NewZealand Digital Library Project at University of Waikato and developed and distributed in cooperation with UNESCO and the Human Info: An NGO. It has been issued as Open-Source, multilingual software under the GNU General Public License.Greenstone not only serves and harvests documents and collections over OAI-PMH but also exports to or imports collections from METS (Metadata Encoding and Transmission Standards).

16

Digital Repository Software: Commercial

Apart from the above Open Source Software, some commercially developed software also available for digital repository. The name of few is mentioned herein below;

CONTENTdm

It is developed by the University of Washington. The software has tools for acquiring or creating collections; tools for storage of the content and a set of tools for displaying and retrieving the objects.

Hyperion

It provides the tools for organizing, storing and accessing to digital files by searching both associated Meta data and full text of text files.

Meta Source

Meta source is a set of tools used to manage digital collections, including, digital object storage, crawling external collections and support for Metadata schemes.

VITAL

VITAL is an institutional repository software developed by VTLS team. VITAL is a set of workflow extensions, management utilities and enhanced searching capabilities build on Fedora Repository Architecture.

17

Statement of the Problem

The Departments in the University of Jaffna are actively involved in publishing articles in magazines, research papers in Journals, presenting papers at conferences, writing book reviews, preparing Cases, compiling edited books, and presenting newspaper articles.

Currently, these articles, research material and scholarly publications – intellectual outputs of Jaffna University exists as disorganized and decentralized digital objects held in personal computers, web servers or storage Medias. They represent a vital part of the scholarly property of the University, and at the same time they are also a part that is the most endangered of being lost.

18
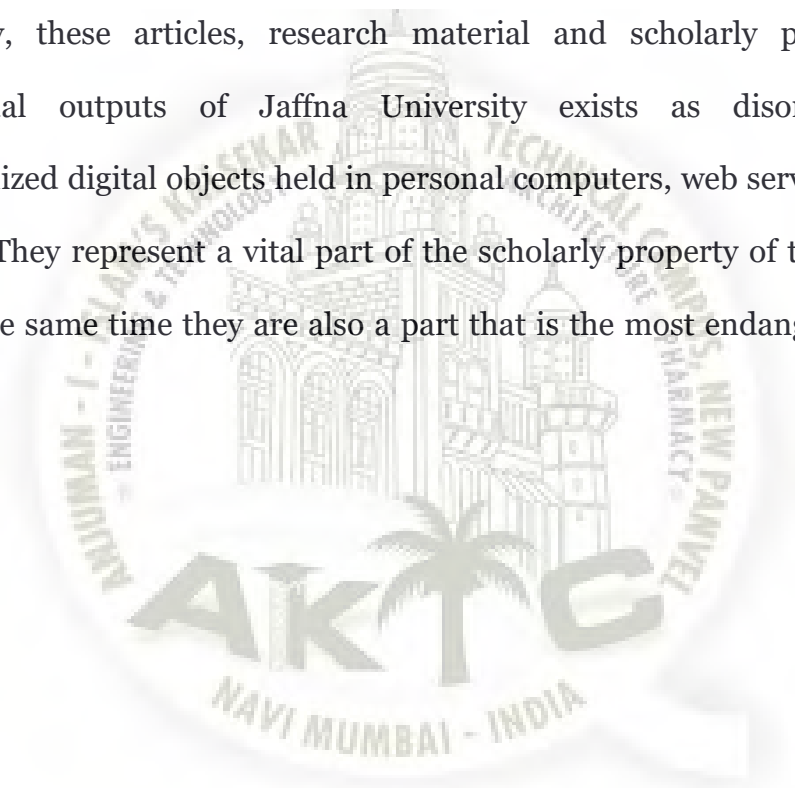
The faculty staff needs to store and retain their intellectual assets. They also need to make their work available and visible to others within and outside the University, while managing their digital rights and maintaining the integrity of their work.

As a Solution, a Digital repository can take the responsibilities to capture and maintain these intellectual outputs. Thus, by this project works, an effort has be previewed to maintain a DSpace Digital Repository System in the University of Jaffna

19

Chapter 2

REVIEW OF LITERATURE

There have been number of Literature published related to the concept of building institutional repositories. Among those, the interestingly referred conceptual works done by each authors of the articles are described in the followings,

1. **Kamila, Kanchan** discussed about the concept of Institutional Repository (IR), its relevance, merits, software requirements and the current trends in India, with special reference to the initiatives at Burdwan University.

2. **Mishra et al.** explained and shared the experiences they have gained from the digitization to accessing ETD, and the challenges they faced, enhancements they incorporated with special emphasis on technical developments, and the lessons they have learnt during the various stages of development of their project for ETD.

3. **Poornima N, Jayashree S and Indrani, V** explained the working model of NAL's Institutional Repository. They discussed the technology employed and methodology adopted in building an institutional repository.

20

4.   The collection process of different datatypes, processing and depositing the same to IR are also discussed in detail.

5. **Prasad, A.R.D** discussed the DSpace digital library software. The implementation of DSpace to build a digital library of Library and Information Science is also explained. He presented a list of features of DSpace to justify its choice over competing open source DL software. He also presents the various collections that are built under the Librarians' Digital Library (LDL). He further discussed the OAI-PMH, handles for identification of digital documents and lucent search engines query language.

6. **Dr. Sharma, P.L** briefed out about Repository, Institutional Repository, their Benefits, and Essential Elements of IR. And Author has also discussed about the software's that are easily available to create and maintain an institutional repositories i.e. Open Source software and Commercial digital repository software.

7. **Sreekumar, M.G et al.** emphasized the need for deploying interoperable open access IRs, and share their experience in creating a state-of-art scholarly Institutional Repository using the DSpace software.

The method of developing the IR, which include the institutional open access policies, the installation, configuration and customization of the software along with the related workflow operations such as defining the communities and collections, content development and management, designing and creating standard metadata sets are also illustrated by the Authors.

8.Dudu Sizakele Nkosi learned about the roles of Academic Libraries with reference to the establishment of an Institutional Repository. Their institutional repository provides opportunity for citations, providing an opportunity for the required expose to other web based access tools. It provided an alternative to the conventional publishing channel. Their paper focused on the practical experience of the University of South Africa library as a result of the implementation of the Institutional Repository through a pilot projects. The paper concluded by highlighting some solutions to ensure that the research work of academics is exposed to the world. It also provided an alternative to start providing access for potential articles which can get into the ―international journals‖.

9.Ghosh, S. B. and Kumar Das, Anup discussed various repository initiatives in India. Some initiatives have also been taken in the area of metadata harvesting services. Gradually, there has been a realization of the usefulness of the open access by various institutions particularly, the public funded ones. They tell the future of open access in India is dependent upon a proper policy and a framework. They also suggested that in the implementation of open access, LIS professionals should play a pro-active role in the growth of collection in the institutional repositories.

10.Hendry, Julia discussed the process of using the institutional repository to manage the University of Illinois Medical Center's electronic policies and procedures documents.

11.K¨orber, Nils and Suleman, Hussein described efforts to highlight usability issues while setting up and configuring DSpace. Article explained about the User evaluations performed on a recent version of DSpace were followed by participatory design of a tool to increase usability by abstracting away the lower-level details. It exhibited that Users agreed that such new tool would be suitably usable. Thus by the article, it was found that significant usability problems exist, but these problems may in fact be easily addressed.

23
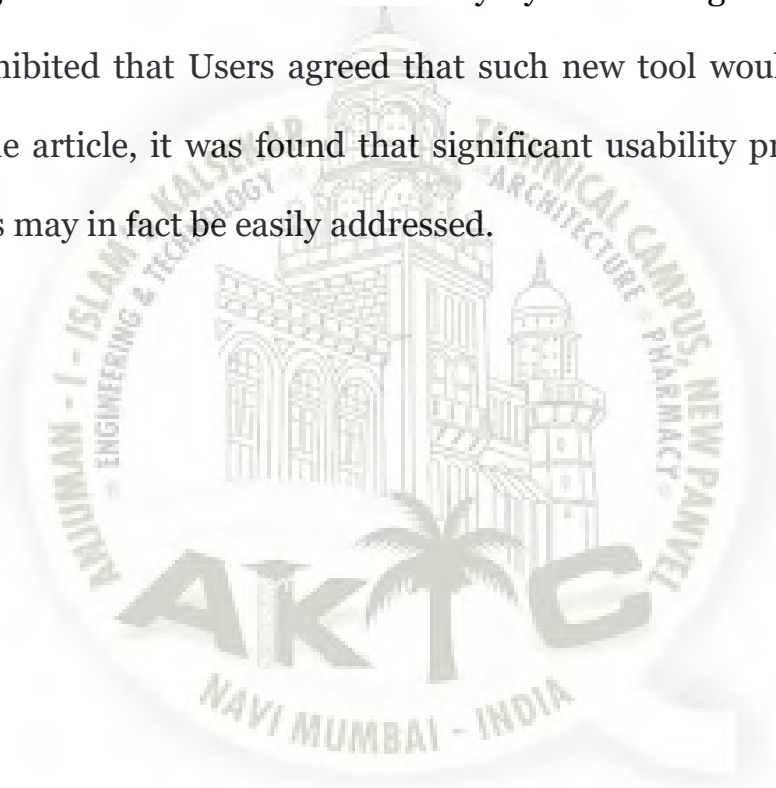
Reference

1. Kamila, Kanchan Institutional Repository Projects in India, 7th International CALIBER-2009, February, 2009.

2. Mishra, R et al. Development of ETD Repository at IITK Library using DSpace, ICSD-2007, pp. 249-259, 2007.

3. Poornima N, Jayashree S and Indrani V Institutional Repository at National Aerospace Laboratories: A Case Study, ICSD-2007, pp. 274- 284, 2007.

4. Prasad, A.R.D. LDL: Librarians' Digital Library, DRTC

5. Dr. Sharma, P.L Institutional Repositories: An Essential tool for Information and knowledge Sharing, National Hydroelectric Power Corporation Limited, Faridabad, 2008

6. Sreekumar, M.G et al. Institutional Repositories for Knowledge Management in Academic and Research Institutions, ICSD-2007, pp. 260-273, 2007.

7. Dudu Sizakele Nkosi, Establishing an Institutional Repository: A Case Study at University of South Africa, Conference on Electronic Publishing and Dissemination, Dakar, Senegal, 6-7 October, 2008.

8. Ghosh, S. B. and Kumar Das, Anup open access and institutional repositories – a developing country perspective: a case study of India, world library and information congress: 72nd ifla general conference and council, 20-24 august 2006, Seoul, Korea.

9. HENDRY, JULIA Managing University of Illinois (at Chicago's University) Medical Center Policies using DSpace, SAA Campus Case Studies, December 2008.

10. K¨orber, Nils and Suleman, Hussein, ICADL 2008, LNCS 5362, pp. 31– 40, 2008. Springer-Verlag Berlin Heidelberg 2008.

25

Chapter 3

INSTITUTIONAL REPOSITORIES

National Level

In the evidence of searching lanka websites, In Sri Lanka, some institutions, like NSF (National Science Foundation), WHO (World Health Organization) have established open access institutional repositories (IRs) using Dspace.

National Science Foundation, Sri Lanka

NSF has the collections in the fields of Agriculture, Biological Sciences, Biochemistry, Genetics, Education, Environmental Science, Fisheries, Medical Science, Meteorology, Natural Resources, Social Sciences, Veterinary Science

26

World Health Organization, Sri Lanka

World Health Organization, Sri Lanka has the communities on Communicable Diseases, Maternal and Child Health, Non-Communicable Diseases, Publications, WHO General etc.

The Home page of WHO Dspace appears as,
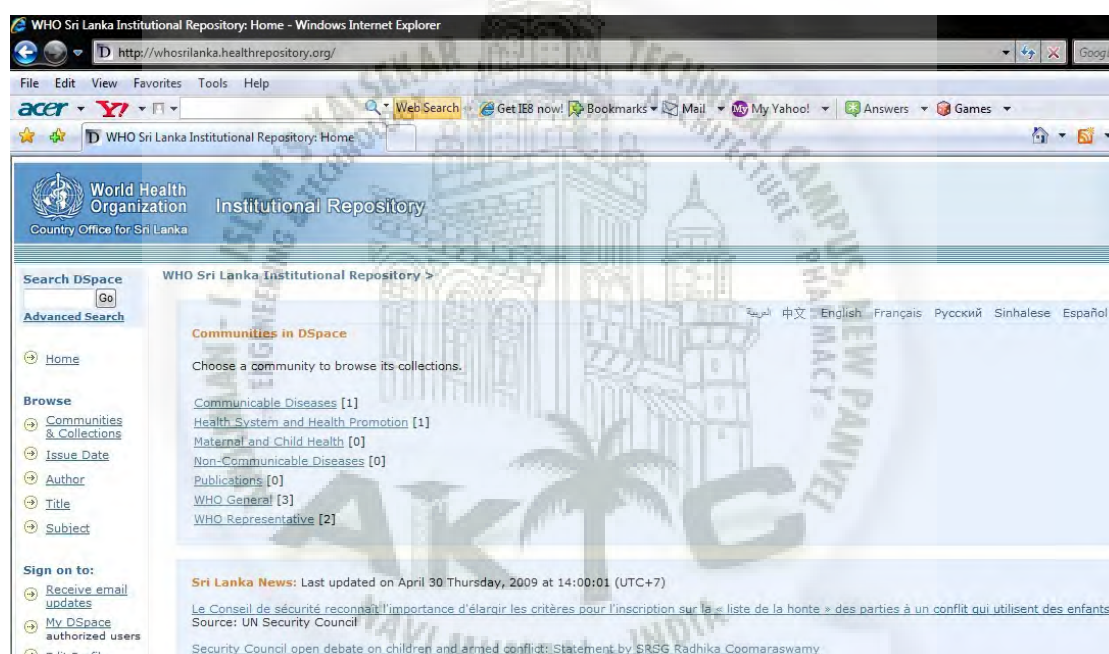
International Level

In India, some institutions, like Indian Institute of Management, Kozhikode; Indian Institute of Technology, Bombay; and Librarian's Digital Library (LDL) of Documentation Research and Training Centre (DRTC), Bangalore are the famous examples of Institutional repositories over the India that disseminates research outputs of respective institution.

27

Indian Institute of Management, Kozhikode (IIMK)

This is a scholarly archiving facility for the IIMK community, using the DSpace software. This service enables the Institute community to archive their preprints, post prints and other scholarly publications. In keeping with the objectives of the Open Access movement, IIMK expects this service to facilitate the Institute researchers in self-archiving and long-term preservation of their scholarly publications, provide easy access to these publications worldwide and improves impact of their research. While dspace@iimk can be accessed by anybody, submission of documents to this archive is limited to the IIMK research community.

The Home page of IIMK Dspace appears as,
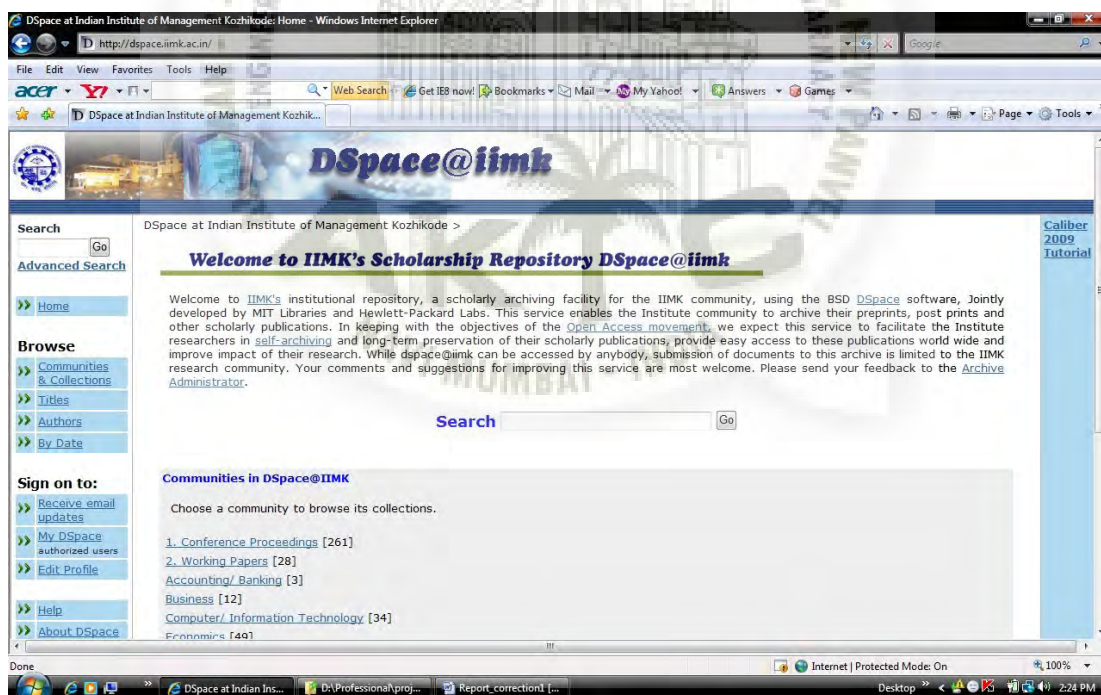
Indian Institute of Technology, Bombay;

Central library, IIT Bombay has set up an open archive of institutional publications including journal articles (preprints/postprints), conference papers, book chapters, reports, theses and dissertations, and other scholarly publications using open source software, DSpace.It enables the Institute community to deposit (self archive) their publications using a web interface, and organizes these publications for easy retrieval. It is expected by IIT Bombay that the repository will evolve as a major source of reference for all IIT Bombay publications accessible on the net.

The Home page of IIT Bombay-Dspace appears as,

Documentation Research and Training Centre (DRTC), Bangalore

The repository accepts scholarly publications from any professional or researcher who belongs to the respective subject. Librarian's Digital Library (LDL) of Documentation Research and Training Centre (DRTC), Bangalore is an example of subject-specific repository for the library and information professionals. We may search or become a member or subscribe to LDL. If we wish to upload our publications in the field of Library and Information Science, we have to write to ard@drtc.isibang.ac.in. To encourage authors in LIS, they have minimal review policy. The only policy requirement is that the publication should be in LIS field.

The Home page of DRTC Dspace appears as
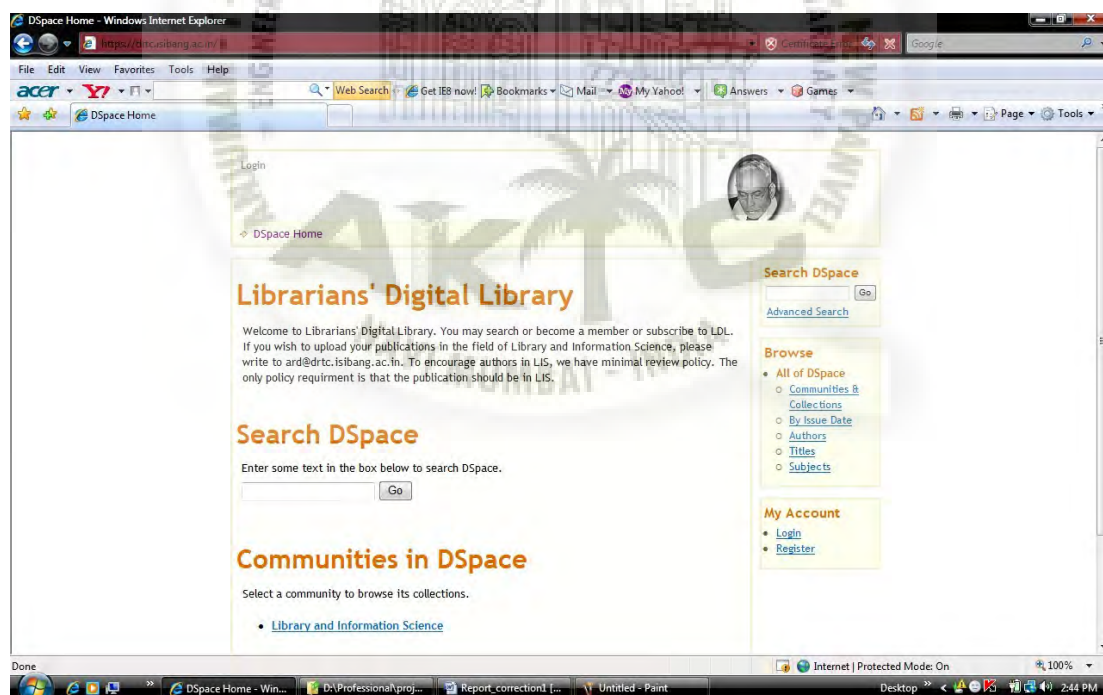
Apart from the described repositories some other repositories of India are providing their services to their societies as example National Institute of Technology, Rourkela; National Aerospace Laboratories, Bangalore; National Chemical Laboratory, Pune; Information and Library Network (INFLIBNET), Ahmadabad; National Institute of Oceanography, Goa; Raman Research Institute, Bangalore; etc. have established open access institutional repositories (IRs) that disseminate research outputs of respective institution. Sometimes, these are self-archived. Otherwise, administrator of the repositories collects the research documents from different sources and submits the documents to the IR on behalf of the persons concerned.

Another band of digital repositories also exist in India that store and provide access to specific subject collections of documents. These repositories accept scholarly publications from any professional or researcher who belongs to the respective subject. Librarian's Digital Library (LDL) of Documentation Research and Training Centre (DRTC), Bangalore (described before) is an example of subject-specific repository for the library and information professionals. Another subject-specific repository established in India is OpenMed@NIC, maintained by National Informatics Centre, New Delhi. OpenMed@NIC stores and provides access to biomedical literature. Other kind of digital repositories existing in India stores and provides access to document type collections. Vidyanidhi of University of Mysore is an example of document type collection that stores and provides access to theses and dissertations. Vidyanidhi accepts any thesis or dissertation from any researcher or student that is accepted in any of the Indian universities or institutions.

The Famous DSpace Repositories other than in India

MIT (Massachusetts, USA)

DSpace@MIT contains MIT Research in digital form, including preprints, technical reports, working papers, theses, conference papers, images, and more.Here the DSpace does not contain "all" MIT's research and is limited to digital research products only.

The Home page of MIT Dspace appears as,
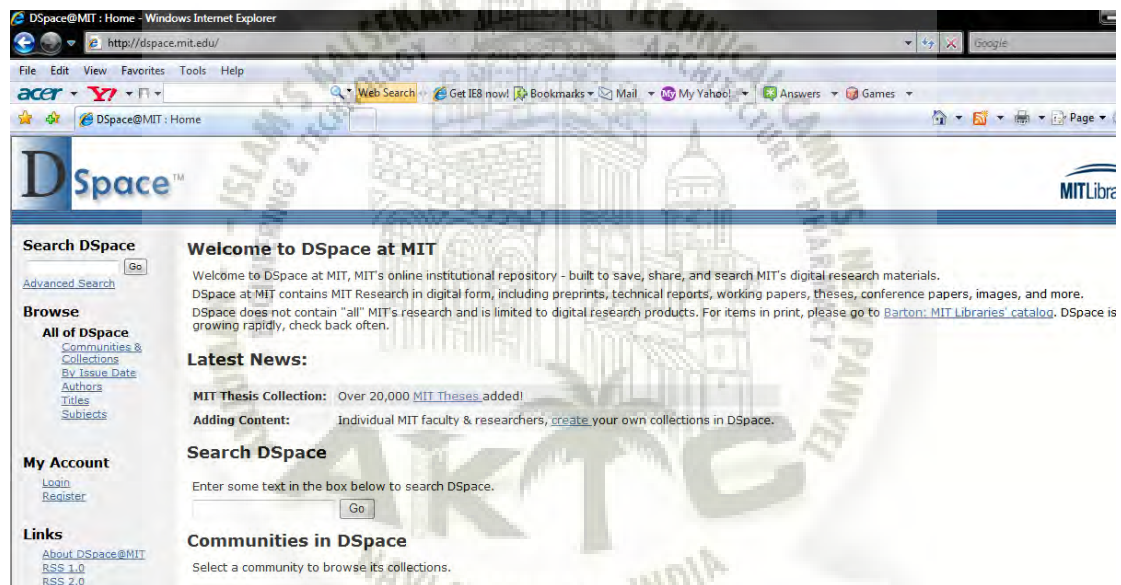
Cambridge University (Cambridge, UK)

DSpace@Cambridge is the institutional repository of the University of Cambridge. The repository was established in 2003 to facilitate the deposit of digital content of a scholarly or heritage nature, allowing employees and their departments at the University to share and preserve this content in a managed environment.The repository has been established to facilitate both dissemination and preservation of digital material created by the members of the University of Cambridge. Its existence demonstrates the University's continued commitment to the stewardship of scholarly material in a digital context, as indeed it has shown this commitment for paper based material for 800years.

The Home page of Cambridge Dspace appears as,

33

Cranfield University (Bedfordshire, UK)

DSpace@ Cranfield University is a digital repository of research output from Cranfield University. It contains a growing subset of research undertaken at Cranfield University including journal pre-prints and peer-reviewed journal reprints, full-text digital Theses and Dissertations, book chapters, working papers and technical reports.The Home page of Cranfield University Dspace appears as,

Drexel University (Pennsylvania, USA)

DSpace@Drexel University is a centralized virtual space to access unique digital resources produced by the Drexel community. Administered by the Drexel Libraries is committed to providing permanent open access to the digital works of Drexel University.The Home page of Cranfield University Dspace appears as,
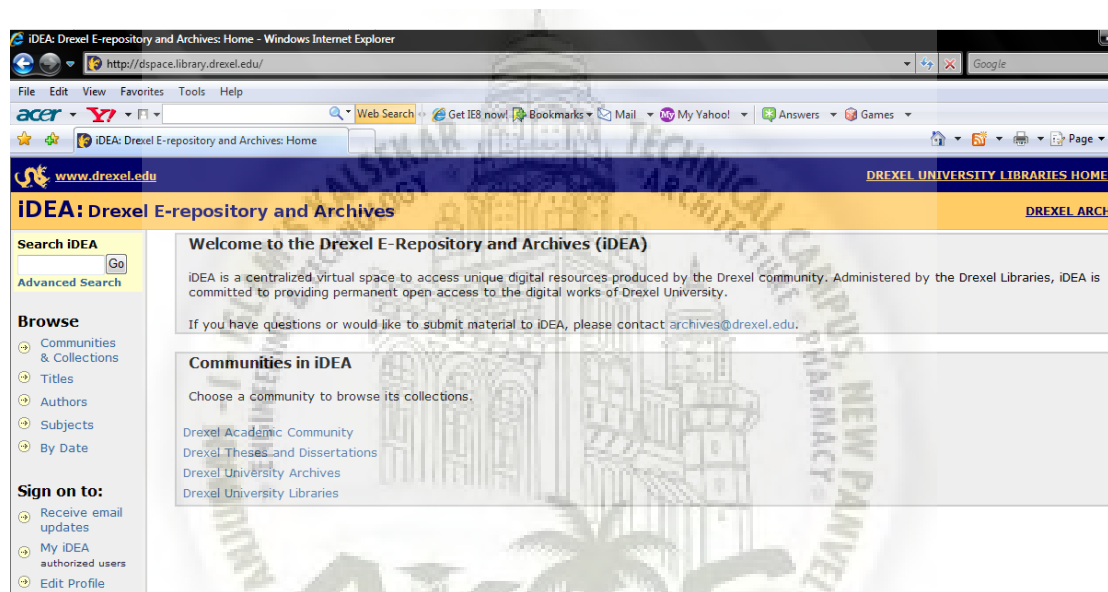
Some other international Dspace institutional repository communities are: Duke University (USA), University of Edinburgh (UK), Erasmus University of Rotterdam (Netherlands), Glasgow University (UK), Hong Kong University of Science & Technology Library, University of Oregon (USA).

35

Chapter 4

RESEARCH DESIGN

Objectives

The design of the project concentrated on four main objectives of the institutional repository at University of Jaffna:

- To create global visibility for research;

- To collect content in a single location;

- To provide open access to institutional research output by self-archiving it;

- To store and preserve other institutional digital assets, including unpublished or otherwise easily lost ("grey") literature (e.g., theses or technical reports).

Here, much care is to be taken on copyright is owned by the author or institute, or for which permission has been obtained to include a copy of the work in the repository.

36

Collections to the Repository:

The collections of University of Jaffna that are needed to be organized are,

- Pre-prints of articles or research reports submitted for Publication

- The text of journal articles accepted for publication

- Revised texts of published work with comments from academic readers
- Conference papers

- Teaching materials

- Student projects

- Doctoral theses and dissertations

- Datasets resulting from research projects

- Committee papers

- Works of art

- Photographs and video recordings

- Official Documents

Selection of Institutional Repository Software: Dspace

The history and specific features of the repository software: Dspace, are described here in order to see its uniqueness in the selection, In March 2000, Hewlett-Packard Company (HP) awarded $1.8 million to the MIT Libraries for an 18-month collaboration to build DSpace, a dynamic repository for the intellectual output in digital formats of multi-disciplinary research organizations. HP Labs and MIT Libraries released the system worldwide on November 4, 2002, under the terms of the BSD open source license, one month after its introduction as a new service of the MIT Libraries. As an open source system,

DSpace is now freely available to other institutions to run as-is, or to modify and extend as they require to meet local needs. From the outset, HP and MIT designed the system to be run by institutions other than MIT, and to support federation among its adopters, in both the technical and the social sense.

So what is DSpace? The DSpace system provides a way to manage these research materials and publications in a professionally maintained repository to give them greater visibility and accessibility over time as just we need for University of Jaffna.

DSpace was built breadth-first: it supports every function that a research organization needs to run a production digital repository service, but as simply as possible. DSpace is designed to make participation by depositors easy. The system's information model is built around the idea of organizational "Communities"—natural sub-units of an institution that have distinctive information management needs.

38

Metadata Representation

Dublin Core metadata element set is preferred for our intuitional repository as much of databases use this Dublin Core Standard, this will enable the university to easily update the repository with collection data received from with Dublin Core Standard. As Dspace is used as our OAI-Repository Software, Dublin Core metadata element set is a standardized vocabulary for describing a variety of our digital documents.DSpace uses a qualified Dublin Core metadata standard for describing items intellectually (specifically, the Libraries Working Group Application Profile). Only three fields are required: title, language, and submission date, all other fields are optional. There are additional fields for document abstracts, keywords, technical metadata and rights metadata, among others. This metadata is displayed in the item record in DSpace, and is indexed for browsing and searching the system (within a collection, across collections, or across Communities). For the Dissemination Information Packages (DIPs) of the OAIS framework, the system currently exports metadata and digital material in a custom XML schema while we work with the METS [14] community to develop the necessary extension schemas for the technical and rights metadata about arbitrary digital formats.

User Interface

User interface of our repository would be web-based as DSpace's current user interface is web-based. There are several interfaces will be in the university repository one for submitters and others involved in the submission process, one for end-users looking for information, and one for system administrators.

The end-user or public interface will support search and retrieval of items by browsing or searching the metadata. Once an item is located in the system, retrieval is accomplished by clicking a link that causes the archived material to be downloaded to the user's web browser. "Web-native" formats (those whichwill display directly in a web browser or with a plug-in) can be viewed immediately just after clicking; others must be saved to the user's local computer and viewed with a separate program available in the local computer that can interpret the file (e.g., a Microsoft Excel spreadsheet).

40
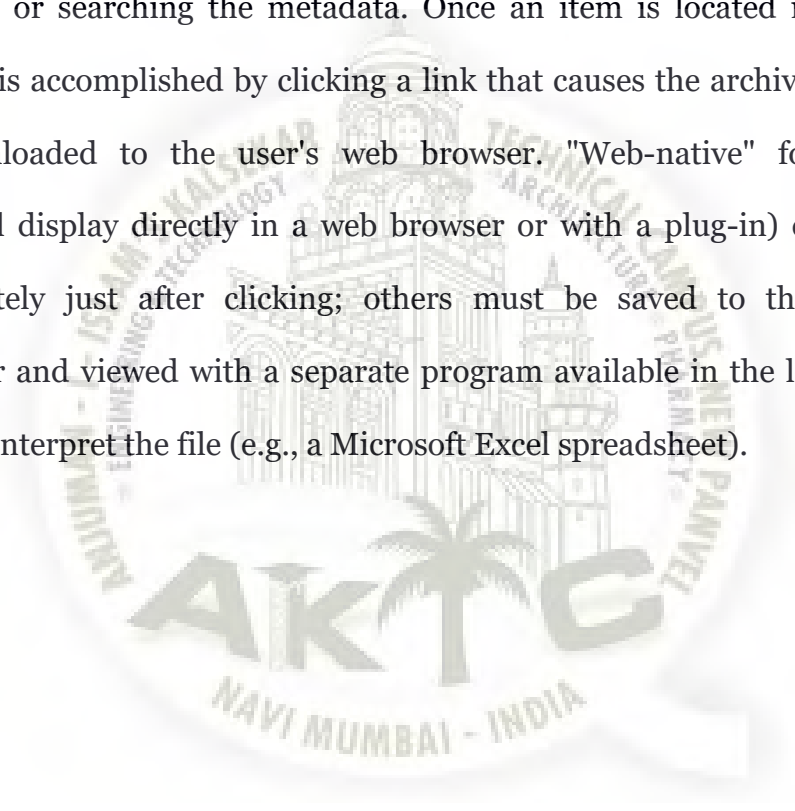
Workflow to the Repository

Multiple workflows have to be available in the repository system at a time. DSpace is tackling the complex problem of how to accommodate the differing submission workflows needed for a multidisciplinary system. In other words, different DSpace Communities, representing different schools, departments, research labs and centers, have very different ideas of how material should be submitted to DSpace, by whom, and with what restrictions. Who is allowed to deposit items? What type of items will they deposit? Who else needs to review, enhance, or approve the submission? To what collections can they deposit material? Who can see the items once deposited? All of these issues are addressed by the Community representatives, working together with the Libraries' DSpace user support staff, and are then modeled in a workflow for each collection to enforce their decisions. The system models "e-people" who have "roles" in the workflow of a particular Community in the context of a given collection. Individuals from the Community are registered with DSpace, and then assigned to appropriate roles.For example, a department may choose to have two collections: one for working papers and another for datasets. They may then decide that any member of the faculty can deposit items to either collection directly, and that any member of the general public can have access to these collections. In this example the workflow is very simple, and the only "role" is that of submitter.

In a more complex example, the same department may have a working paper collection that requires tight editorial control by the head of the department. In this case, they may choose to again designate a departmental staff as "submitters", but also designate a small group of people as "reviewers", an administrative staff person as a "metadata editor", and the head of the department as the final "coordinator". An item deposited by a faculty member would then go through a process of review, cleanup and approval before finally being deposited to the relevant DSpace collection. Each person with a role to play in this process is notified of the new submission, and goes to a personal workspace in the system to perform their assigned task. Items that do not make it through the process are not archived in the system.
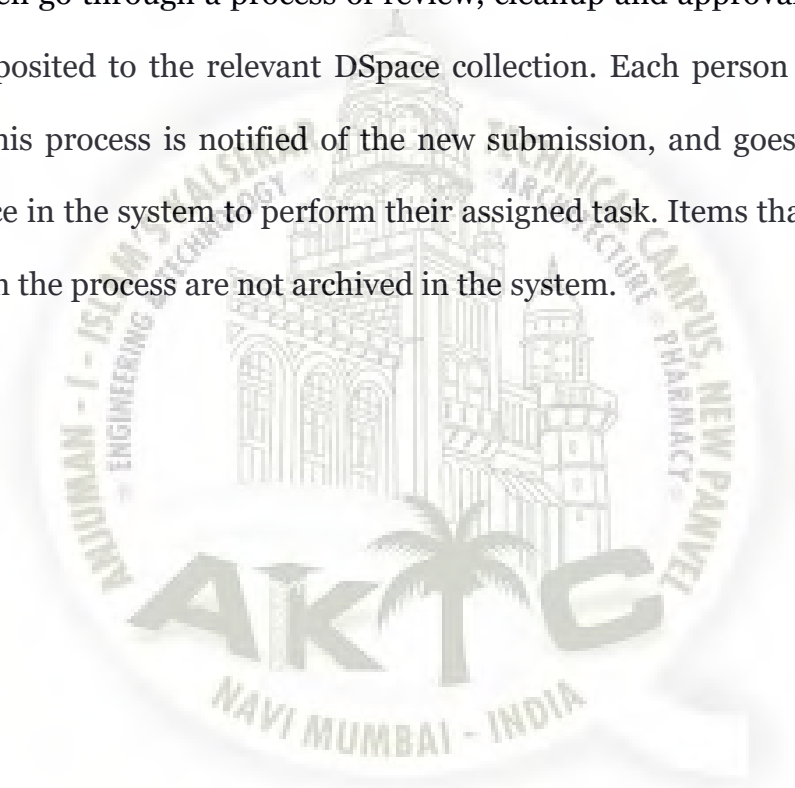
Technology Platform of the Repository

The university repository system needs to be run on the UNIX and windows XP/2000 platform, and comprises other open source middleware and tools, and programs written by the DSpace team. All original code would be in the Java programming language. Other pieces of the technology stack will include a relational database management system (PostgreSQL), a Web server and Java servlet engine (Apache and Tomcat, both from the Apache Foundation), etc.

The university must have adequate resources to use the dspace system, including adequate hardware that runs the UNIX (or Windows) operating system with enough storage, and a UNIX (or Windows) systems administrator with enough skills to install configure, and manipulate the repository system. The University will also want the services of a Java programmer who can localize and customize for them, or enhance it, although this is not absolutely necessary to initially run the system.

System Architecture of Dspace

The repository architecture would be straightforward three-layer of architecture, including storage, business, and application layers, each with a documented API to allow for future customization and enhancement. The storage layer is implemented using the file system, as managed by PostgreSQL database tables. The business layer is where the DSpace-specific functionality resides, including the 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. Finally, the application layer covers the interfaces to the system: the web UI and batch loader, in particular, but also the OAI support and Handle server for resolving persistent identifiers to DSpace items. This is the layer that will get much of the attention in future releases, and define Federation services across the range of institutions adopting DSpace.

Why DSpace is selected?

As seen in the above descriptions, the reasons why we chose DSpace are:

1. Dspace is corded as open and a basis platform to build an Institutional Repository and its collections which can be easily searchable and retrievable by the Web with appropriate system architectures.

2. Dspace is an open source technology platform which can be easily customized or extend its capabilities in the Windows or UNIX platforms.

3. Dspace is a free service model for open access and/or digital archiving for perpetual access with Dublin core metadata with many options on metadata elements.

4. To make available institution-based scholarly material in digital formats.

   The collections will be open and interoperable with multiple workflows and different interfaces.

45

Tools for Building IR

Hardware

For building an institutional repository we need some specific features for Hardware elements,

For example,

I.   HP Server rx2600, powered by dual 64-bit Intel Itanium 2 processors (900MHz), 2GB RAM 26 GB internal disk storage. Total capacity can be six terabytes.

II.   Sun Fire 280R Server, two 900MHz UltraSPARC-III Cu processors, 8MB E-cache, 2GB memory, two 36GB 10,000rpm HH internal FCAL disk drives, DVD, 436-GB, or 12 x 26.4 Gbyte 10K RPM disks, Sun StorEdge A1000 rackmountable w/ 1 HW RAID controller, 24MB std cache.

III. Dell PowerEdge 2650 with dual Xeon processors (2.4GHz), 2GB RAM 2 x 73 GB scsi disks. One 2.5TB Apple XServe.

The above Hardware themes is well suit for building an repository for an institutional purpose of big storage as University of Jaffna needs, but for this project work for building a sample repository the available Intel Core Duo CPU @ 2.20 GHz of 2GB Ram with 250 GB HDD storage can be used.

Software:

The assembly of software needed for building the repository is,

• Java SDK 1.4.2

• Apache 2.0.54

• Tomcat 5.0.28

• Apache Ant 1.6.5

• PostgreSQL 8.0.2

• DSpace 1.4 alpha 1

The above list of software is to be installed in a sequence order following strict formalities as described in the methodology for successfully run the Dspace system.

Operating System needed to install the collection of software in window platform is Windows XP or Windows 2000. Drive Partition needed is NTFS.

METHODOLOGY

The concepts of institutional repositories were studied for having much background on it. Then plan was prepared to collect and set up the tools for building the repository and how to get the necessary article that match with University of Jaffna for uploading to my sample repository.

As a starting point towards building the repository, the Windows Operating System – XP was installed with NTFS partitions in an Acer Laptop System of Intel Core Duo CPU @ 2.20 GHz of 2GB Ram with 250 GB HDD storage.

The assembly of software needed for building the repository, Java SDK 1.4.2, Apache 2.0.54, Tomcat 5.0.28, Apache Ant 1.6.5, PostgreSQL 8.0.2, DSpace 1.4 alpha 1 were collected from a Data Source available in the Department of Library and Information Science, Bharathidasan University and installed in sequence as follows,

48

Installing Dspace on Windows

Sequence of Software installed

Installing Java 2 SDK

- j2sdk-1_4_2_09-windows-i586-p.exe was run

- Installation was finished at C: \ forming C:\j2sdk1.4.2_09

- PATH: C:\j2sdk1.4.2_09\bin was added

- Environmental Variable was set as JAVA_HOME=C:\j2sdk1.4.2_09

Installing Apache 2.0.54

- A Folder -www‖ was created on C:\

- The file apache_2.0.54-win32-x86-no_ssl.msi was run

- Apache HTTP Server 2.0 was installed to folder C:\www

49

Installing Tomcat 5 and mod_jk2

- Tomcat5 folder was copied from data source to the directory C:\www

- The file mod_jk2.so was copied from data source and pasted under folder C:\www\Apache2\modules\

- The file tools.jar was copied from C:\j2sdk1.4.2_09\lib\ to

- C:\www\Tomcat5\common\lib\

- The file workers2.properties was copied from data source and placed under folder C:\www\Apache2\conf\

- The file C:\www\Apache2\conf\httpd.conf was edited and following code was inserted at last.

    LoadModule jk2_module modules/mod_jk2.so

    JkSet config.file
    "C:/www/Apache2/conf/workers2.properties"

- The file C:\www\Tomcat5\conf\jk2.properties was edited and following

- Line was inserted at last.*channelSocket.port=8009*

Note: it was confirmed that Apache has been started at this point

- -C:\www\Tomcat5\bin>service.bat install‖ was run at command prompt

- Now Apache Tomcat Service was started.

- Tomcat server at <u>http://localhost:8080/</u> was test run.

50

Installing Apache Ant 1.6.5

- The Folder apache-ant-1.6.5 was pasted to c:\

- PATH : C:\apache-ant-1.6.5\bin was added

- Environmental Variable was set as ANT_HOME= C:\apache-ant-1.6.5

Installing PostgresSQL8.0.2

- Postgresql-8.0.msi was run

- Service Configuration:

  A user for operating Postgre service was created.

  User name: postgre ; Password: postgre

- Initializing Database Cluster (at this stage we need to start a service ―secondary logon service‖):

- An Internal Database user was created

- User name: dspace; Password: dspace

- The installation was finished

- Then, Start>Programs>PostgreSQL 8.0>Pg Admin III

- Then a Right click was made on -PostgreSQL Database Server 8.0‖ and a click was executed on ―Connect― (PW: dspace)

- PostgreSQL Database Server 8.0 was opened, then -Database‖ was selected. After right clicking on -Database‖, New object >> New Database was opened.

- A database -dspace‖ was created as owned by the user -dspace‖ with UNICODE encoding

51

## Installing DSpace 1.4 alpha 1

- The Folder Dspace-1.4 alpha 1 available from data-source was

  pasted to C:\

- The file postgresql-8.0-310.jdbc2.jar   was   copied   from

  C:\Program Files\PostgreSQL\8.0\jdbc\ to C:\ Dspace-1.4

  alpha1\lib.

- PostgreSQL service was ensured running

- The paths in C:\ Dspace-1.4 alpha 1\config\dspace.cfg. was updated.

- A Folder -Dspace‖ was created on C:\

- C:\Dspace-1.4 alpha 1\ant fresh_install was run

- An administrator account was created by running

  C:\dspace\bin\dsrun org.dspace.administer.CreateAdministrator

- The search and browse indices were initialized by running

  both, C:\dspace\bin\dsrun

  org.dspace.browse.InitializeBrowseC:\dspace\bin\dsrun

  org.dspace.search.DSIndexer

- The .war Web application files from C:\ Dspace-1.4 alpha1\build\

  were copied to C:\www\Tomcat 5.5\webapps

  At this point Tomcat service was stared.
- By browsing http://localhost:8080/dspace   DSpace home

  page   was successfully seen with a grace.

52

DSpace Configuration & Customization

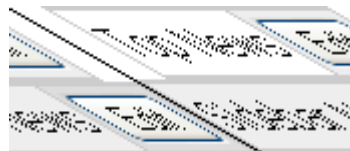## Top News & Sidebar News

After logging in to my Dspace   Step 2
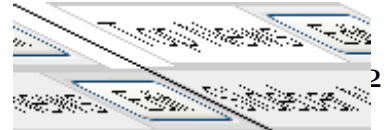
Step 1

Customizing the Web User Interface

Header Image of named Dspace-blue.gif is designed as desired for the

UniversityofJaffna,   which is C:\www\Tomcat5\webapps\dspace\image\

Removing the "About DSpace Software" link from Header

The file header-default.jsp at C:\www\Tomcat5\webapps\dspace\layout\

was edited at <%-- DSpace logo --%> and the cell was customized as required

1. Giving a "Welcome!" Message in the Home Page The

file home.jsp at C:\www\Tomcat5\webapps\dspace\ was edited to establish our

welcome message

53

Chapter 5
BUILDING IR

Under this, the major steps followed for building up collection are described with two stems, Community Creation and Collection Building

Community Creation

There are two options of Communities, first one: Top-Level Community, following one: Sub-Community

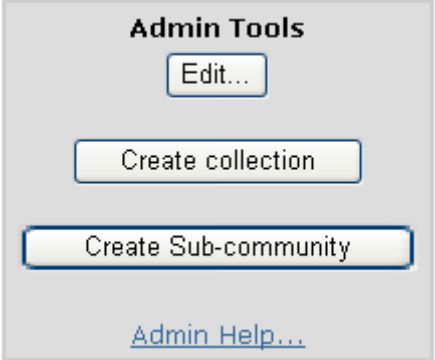### Creating Top Level Community
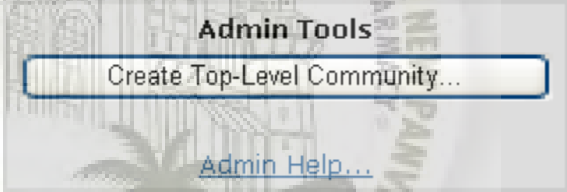
After log in to My Dspace,

Step1

Browse

- Communities & Collections
- Titles
- Authors
- Subjects
- By Date

Step 2

Admin Tools

Create Top-Level Community...

Admin Help...

Creating Sub-Community

Creating Sub-Community is similar to Top Level Creation, but the step 2 becomes as below after finishing Top-Level Community creation.

Admin Tools

Edit...

Create collection

Create Sub-community

Admin Help...

Step 2

Step 3

Name:	Scholar Publication
Short Description	<%if needed only%>

Introductory text (HTML):
```
<html> <%if needed only%>
</html>
```

Copyright text (plain text):
```
<%about copyright of this toplevel community%>
```

Side bar text (HTML):
```
<html> <%if needed only%>
</html>
```

Logo: Upload a logo...

Community's Authorizations: Edit...

Create   Cancel

Collection Building

Admin Tools

Edit...

Create collection

Create Sub-community

Admin Help...

55

Describe the Collection

Please check the boxes next to the statements that apply to the collection.

- ☑ New items should be publicly readable
- ☑ Some users will be able to submit to this collection
- ☐ The submission workflow will include an *accept/reject* step
- ☐ The submission workflow will include an *accept/reject/edit metadata* step
- ☐ The submission workflow will include an *edit metadata* step
- ☐ This collection will have delegated collection administrators
- ☐ New submissions will have some metadata already filled out with defaults

Next >

Step 2

This step is similar to the step3 in Top Level and Sub-Community creation as described before.

Step 3

Authorization to Submit

Who has permission to submit new items to this collection?

You can change this later using the relevant sections of the DSpace admin UI.

Click on the 'Select E-people' button to choose e-people to add to the list.

Click on the 'Select Groups' button to choose groups to add to the list.

| Select E-people... | Remove Selected | | Select Groups... | Remove Selected |

Next >

56

Step 4
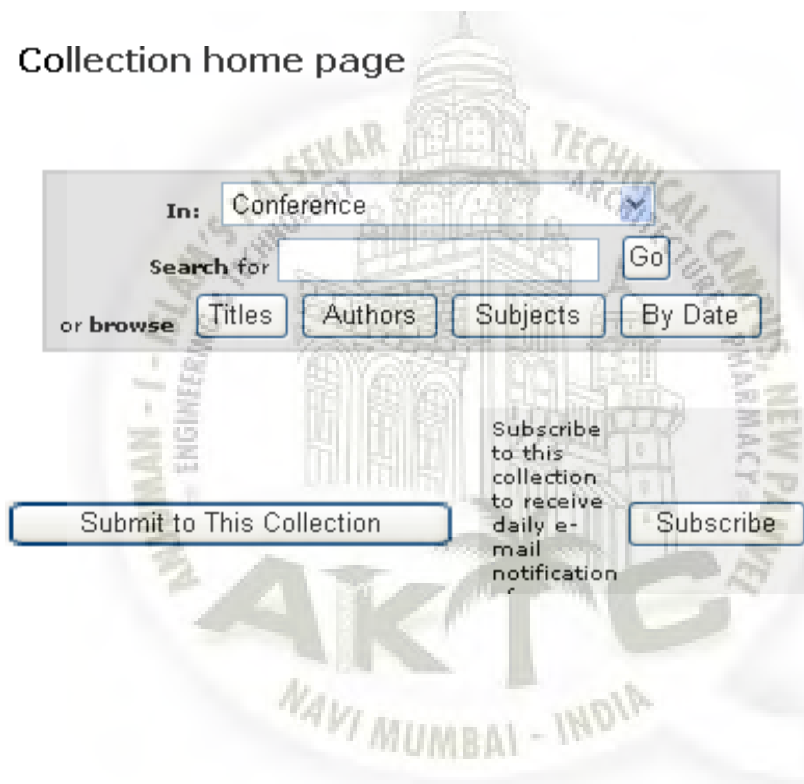
This step is similar to the step3 in Top Level and Sub-Community

creation, but finally a click must be made on -update‖.

Step 5

Assume the collection name that we created is –conference‖

Step 6

Step 7

Step 8

58

Step 9

Describe ▸ Describe ▸ Describe ▸ Upload ▸ Verify ▸ License ▸ Complete

Submit: Describe Your Item

Please fill further information about your submission below.

Enter appropriate subject keywords or phrases below.

Subject Keywords   Repositories   Image Magick   [Add More]

Enter the abstract of the item below.

Abstract   This paper discusses about the concept of Institutional Repository (IR), its relevance, merits,software requirements and the current trends in India, with special

Enter the names of any sponsors and/or funding codes in the box below.

Sponsors

Enter any other description or comments in this box.

Description

[< Previous] [Next >]   [Cancel/Save]

Step 10

Describe ▸ Describe ▸ Describe ▸ Upload ▸ Verify ▸ License ▸ Complete

Submit: Upload a File

Please enter the name of the file on your local hard drive corresponding to your item. If you click "Browse.. a new window will appear in which you can locate and select the file on your local hard drive.

Netscape users please note: By default, the window brought up by clicking "Browse..." will only display files of type HTML the file you are uploading isn't an HTML file, you will need to select the option to display files of other types.

Please also note that the DSpace system is able to preserve the content of certain types of files better than other types.

Document File: C:\Documents and Settings\Ketheeswaren\Des [Browse...]

[< Previous] [Next >]   [Cancel/Save]

59

Step 11

Describe Describe Describe **Upload** Verify License Complete

Submit: File Uploaded Successfully

Your file was successfully uploaded.

Here are the details of the file you have uploaded. Please check the details before going to the next step

More Help...

File	Size	File Format
17[1].pdf	165,676 bytes	Adobe PDF (known)

[Click here if this is the wrong format]

[Click here if this is the wrong file]

You can verify that the file has been uploaded correctly by:

- Clicking on the filename above. This will download the file in a new browser window, so that you can check the contents.
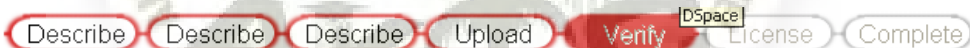- The system can calculate a checksum you can verify. Click here for more information. [Show checksums]

[< Previous] [Next >]   [Cancel/Save]

Step 12

Describe Describe Describe Upload **Verify** License Complete

Submit: Verify Submission

Not quite there yet, but nearly!

Please spend a few minutes to examine what you've just submitted below. If anything is wrong, please go back and correct it by using the buttons next to the error, or by clicking on the progress bar at the top of page.

More Help...

If everything is OK, please click the "Next" button at the bottom of the page.

60

Step 13

Describe ▸ Describe ▸ Describe ▸ Upload ▸ Verify ▸ **License** ▸ Complete

There is one last step: In order for DSpace to reproduce, translate and distribute your submission worldwide agreement to the following terms is necessary. Please take a moment to read the terms of this license, and on one of the buttons at the bottom of the page. By clicking on the "Grant License" button, you indicate th grant the following terms of the license.

More Help...

Not granting the license will not delete your submission. Your item will remain in your "My DSpace" pag can then either remove the submission from the system, or agree to the license later once any queries you have are resolved.

I Grant the License

I Do Not Grant the License

Submission Completes

Describe ▸ Describe ▸ Describe ▸ Upload ▸ Verify ▸ License ▸ **Complete**

Submit: Submission Complete!

Your submission will now go through the workflow process designated for the collection to which you are submitting. You will receive e-mail notification as soon as your submission has become a part of the coll or if for some reason there is a problem with your submission. You can also check on the status of your submission by going to the My DSpace page.

Go to My DSpace

Communities and Collections

Submit to This Collection

61

Chapter 6

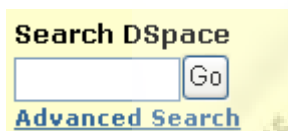# TESTING THE REPOSITORY SERVICE

Home Page

Searching and Browsing

The searching and browsing are tested separately as described in the following
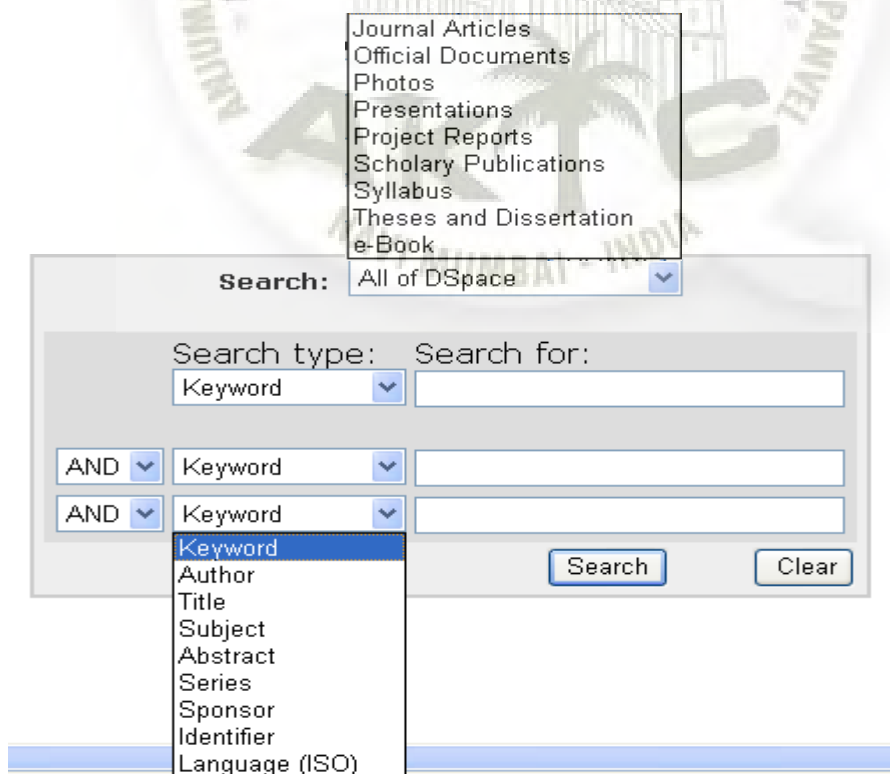
Searching

Under Searching the formal operation on Simple Search and Advanced Search are noted down here,

Simple Search

The operation on this searching box is very common to all.

Advanced Search

63

Example of an Advanced Searching

Consider the following searching environment,

The result for this searching appears as

64

The item description for this searching becomes,

Title: Scientometric Dimensions of Knowledge Management Research in India: A Study based on Scopus data

Authors: Surulinathi, Muthuraj

Keywords: Scientometrics
Knowledge Management

Issue Date: 25-Jan-2009

Abstract: This paper attempts to analyse quantitatively the growth and development of Knowledge Management Research in India in terms of publication output as reflected in Scopus database. During 1999-2007 a t 51 papers were published by the Knowledge Management researchers to various domains: Business, Management and Accounting (24), Engineering (18), Social Sciences (7), Computer Science (6), Decisio Sciences (6), Multidisciplinary (4), Mathematics (3), Environmental Science (2), Agricultural and Biologi Sciences (1), Earth and Planetary Sciences (1), Economics, Econometrics and Finance (1), Materials Sci (1). Year-wise growth of publications and input of records to Scopus database by India is analyzed. Mc than 80 percent of publications were published in journals. The most preferred journals were Internatic Journal of Information Technology and Management (7), Journal of Scientific and Industrial Research (4 Electronic Library (2), Human Systems Management (2), Journal of Knowledge Management (2), Interna Journal of Technology Management (2), International Conference on Information and Knowledge Manag Proceedings (2), Proceedings Frontiers in Education Conference (2), others journals each (1). There wei many as 24 (47.06) papers contributed by single author. There were five authors contributions is zero : more than 5 authors contribution is less than 1(1.96). The publication behavior indicates that the Knov management researchers were lowly selective in publishing.

URI: http://hdl.handle.net/123456789/101

Appears in Collections: Scientometrics

Files in This Item:

File	Description	Size	Format	
Scientometrics-Surulinathi.pdf		678Kb	Adobe PDF	View/Open

The Full Text for this searching is seen as below,

Browsing

Browse

→ Communities & Collections

⊛ Titles

⊛ Authors

⊛ Subjects

⊛ By Date

Say "Browsed by Titles" then the window appears as,

Browse by Title

Jump to: 0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

or enter first few letters: [   ] Go!

Showing items 1-21 of 24.

Next page

Issue Date	Title	Author(s)
8-Apr-2009	APPLIATION FOR APPOINTMENT OF VISITING LECTURER	University of Jaffna
8-Apr-2009	Application for Aptitude Test for BFA Music, BFA Dance, BFA Art & Design	University of Jaffna
9-Apr-2009	Broadband Bible : Desktop Edition	Gaskin, James E.
8-Apr-2009	The Cost of 'Peace' and the Dividends of Terror:Sri Lanka's Nordic Winter?	UTHR
9-Apr-2009	Declaration	University of Jaffna

Say "Browsed by Authors" then the window appears as,

Browse by Author

Jump to: 0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

or enter first few letters: [   ] Go!

Showing authors 1-15 of 15.

DEPARTMENT OF BUILDINGS
Dept of Library & Information Science
Dept. of Physics
Foster, Lonnon R.
Gaskin, James E.
GRE Board
Kenaul, Richard
Ketheeswaren, S
Library
Ranganathan, S.R.
Roselinemary, S

66

Chapter 7

CONCLUSION

The project on building an institutional repository for the university of Jaffna is an attempt to preserve and disseminate different collections available as bulk from various time for achieving a unified way of presentation and searching the digital collection of the university. Further, the project work is expected to master the team of Building the recognized Official Digital Repository to the University of Jaffna in future.

Regarding Dsapce, it is seen as a powerful software for building a repository by its features of Metadata Representation (we can add or remove metadata), User friendly Interfaces, simplest Workflow, well featured Technology platform, and well Organized System Architecture.

Moreover, Dspace is an open source technology platform which can be customized or extended for its capabilities and is a service model for open access and a digital archiving for perpetual access. And the Collections of the Dspace Repository are searchable and retrievable by the Web.

Usually the potential users of an Institutional Repositories in digital environment are countless. Thus, Institutional Repositories content does not need to always_sit' in the Institutional Repositories itself. It would be brought into new digital environments for the University of Jaffna where its scholar contents can be easily consulted, represented and integrated with other current research and educational activities.

This Institutional Repositories idea would be a powerful one that will serve as an engine more broadly for the scholarly enterprises of the University. It is vital that any institutions must recognize Institutional Repositories as a serious and long- lasting commitment to the community of Institution. As such, the University of Jaffna will also have the same commitment at all.

68
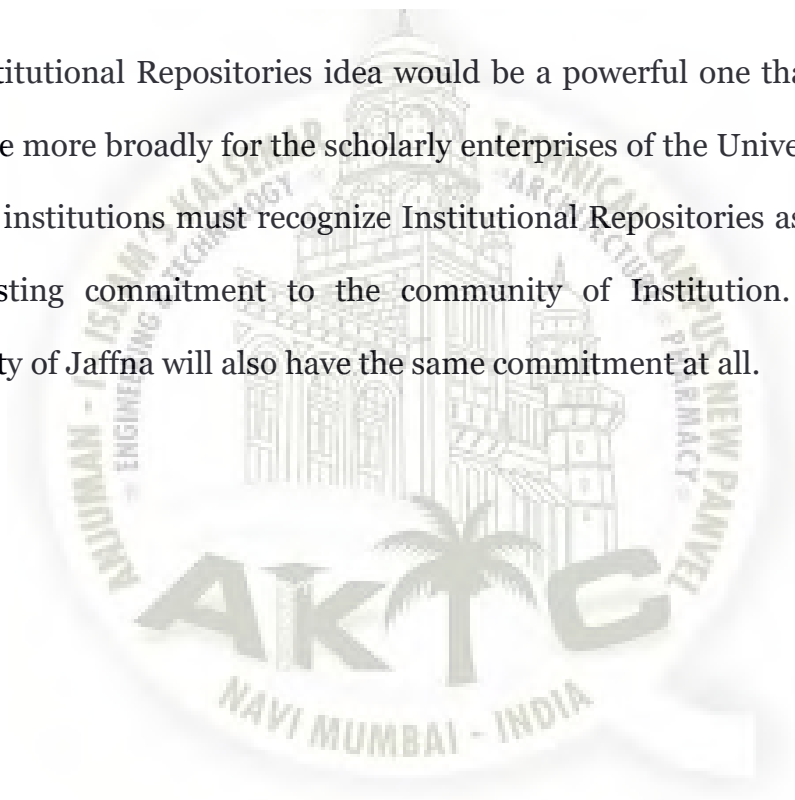
Chapter 8

BIBLIOGRAPHY

1. Kamila, Kanchan Institutional Repository Projects in India, 7th International CALIBER-2009, February, 2009.

2. Mishra, R et al. Development of ETD Repository at IITK Library using DSpace, ICSD-2007, pp. 249-259, 2007.

3. Poornima N, Jayashree S and Indrani V Institutional Repository at National Aerospace Laboratories: A Case Study, ICSD-2007, pp. 274- 284, 2007.

4. Prasad, A.R.D. LDL: Librarians' Digital Library, DRTC

5. Dr. Sharma, P.L Institutional Repositories: An Essential tool for Information and knowledge Sharing, National Hydroelectric Power Corporation Limited, Faridabad, 2008

6. Sreekumar, M.G et al. Institutional Repositories for Knowledge Management in Academic and Research Institutions, ICSD-2007, pp. 260-273, 2007.

7. Dudu Sizakele Nkosi, Establishing an Institutional Repository: A Case Study at University of South Africa, Conference on Electronic Publishing and Dissemination, Dakar, Senegal, 6-7 October, 2008.

8. Ghosh, s. b. and kumar das, anup open access and institutional repositories – a developing country perspective: a case study of India,world library and information congress: 72nd ifla general conference and council, 20-24 august 2006, Seoul, Korea.

9. HENDRY, JULIA Managing University of Illinois (at Chicago's University) Medical Center Policies using DSpace, SAA Campus Case Studies, December 2008.

10. K¨orber, Nils and Suleman, Hussein, ICADL 2008, LNCS 5362, pp. 31– 40, 2008. Springer-Verlag Berlin Heidelberg 2008

11. National Science Foundation of Sri Lanka at http://www.nsf.ac.lk/nslrc/alert.doc

12. Library Site in the University of Drexel at http://idea.library.drexel.edu/

13. Dspace Home Page at http://www.dspace.org/

14. http://www.dlib.org/dlib/january03/smith/01smith.html#3