

ETD- A Scholarly Open Access Institutional Repository of IISc: A case study

**K.Nirmala Devi and S.Venkadesan,
JRD Tata Memorial Library, Indian Institute of Science, Bangalore,
India.**

**Filbert Minj and A.N. Manimalar
NCSI, SERC, Indian Institute of Science, Bangalore, India.**

The IISc, Bangalore was started in 1909 and is considered one among the prestigious institute of its kind in India today. Since then it has grown into a premier institution for research and advanced instruction, with more than 2000 active researchers working in almost all frontier areas of science and technology. It has a very high international standing in the academic world as well. Being an outstanding premier institution in the nation, it aimed to develop digital library for collecting the documents and manage networked information services for the benefit of faculty, students, researchers and other academic community in their education and research. Moreover the quality of an academic institution is reflected by the quality of its student's intellectual products. Thesis is a document that explains the results of the research or scholarship, research process of a PhD student. Indian Institute of Science (IISc) has more than forty departments where active research is going on. The outcome of the research includes theses of PhD, M.Sc (Engg.) and M.E students in Science and Engineering. The rapid and dynamic growth in digital technology is making libraries to convert the useful and valuable resources to digital form and helps to preserve the original materials in electronic form for future use. In this regard, the ETD project started to give greater unlocking scholarly access to the valuable resource for everyone at any time or any place. Instead of knocking at every door to reveal the prestige of an institute, ETD is the one and only hit to make the whole world knowing the esteem of an Institute. Based on the value, use and legal issues PhD and M.Sc (Engg.) theses has been selected for Digitization. With access to ETD, the institute students will be able to find the full texts of related work easily, to read literature reviews prepared by their peers, and to follow hypertext links to relevant data and findings. Also students are taught to be electronic publishers, preparing them for their future work. If they can publish electronically and add to digital libraries, future works will not have to be scanned or rekeyed.

In this paper we discussed the entire digitization process, starting from scanning till the e-theses is made available on ETD archive using D-Space - a open source repository software. Preparation of metadata, browsing features of ETD system and future aim to evolve as a National Repository of Theses and Dissertation also is discussed.

Keywords: ETD, Ph.D, Digitization, Institutional repository, IISc, Theses

1. Introduction

IISc, Bangalore has been at the forefront of research and education in Science and Engineering. Today the Institute has a very high international reputation in the academic world, in view of the fact, it has expertised both in conventional and emerging areas of science and engineering and provides facilities for the post graduate research and course work. The Institute comprises forty and odd departments and centers staffed by high caliber scientists. The institute offers opportunity for young researchers to pursue doctoral and post doctoral studies in science and engineering.

The research scholars are always strived to live up to the expectation of International standards. Hence it is believed that fundamental research is depending upon the library's collection and facilities. In this view, JRD Tata Memorial Library, one of the best scientific and technical libraries in India maintained

with a collection of 5 lakhs books, periodicals, technical reports, standards and patents. It receives over 1734 current periodicals.

To extend the facilities for the students, JRD Tata Memorial Library Annex and Digital Library has been started to initiate the digitization of Institute theses and rare book collection for open access.

For most researchers the theses or dissertations are the foremost work of scholarship they produced. To make this work more readily available to other researchers, as well as to save money, JRD Tata Library and NCSI initiated etd@iisc. ETD is defined as those theses and dissertations submitted, archived or accessed primarily in electronic formats. In this paper we briefly look at the move toward making theses and dissertations available in electronic formats and discuss some of the proposals that have been advanced for dealing production, storage, and dissemination of those works.

2 .The Move to ETD

2.1 Selection of Materials

Traditional methods of archiving and storing theses and dissertations are inefficient and unwieldy. Many theses and dissertations are remain unutilized due to lack of access and availability, with no efficient way for researchers to locate the information that may be contained in them.

Many libraries are now in the process of digitizing information in an effort to preserve it and to make it more widely available. The JRD Tata Library plans to digitize around 6000 theses of science and engineering.

The Indian Institute of Science has more than forty departments, which comes under six divisions namely, Division of Biological Sciences, Chemical Sciences, Electrical Sciences, Information Sciences, Mechanical Sciences and Physical and Mathematical Sciences. The prestige of the institute depends on its research output. The number of PhDs and M.Sc (Engg) in Science and Engineering generated from the Institute would be around 200 annually. Based on the value, use and legal issues, Institute PhD M.Sc (Engg) theses have been selected for digitization.

Good selection technique ensure that resources are invested wisely in digitizing the most significant and useful collection at the lowest possible cost without placing the Institution at legal risk. Poor selection leads to the digitization of materials that are unusable or little value.

2.2 Significant of the Project

Archiving theses and dissertations electronically can help to alleviate some of the problems involved in storage, and making full-text versions available either on the Web would make access almost immediate. Electronic versions on disk, CD-ROM, or other digital electronic media could be cheaper as well.

JRD Tata Library encourages individual scholars to publish their own works on the Web, allowing free access for full text searching. That would allow researchers to be sure that the documents they order or download actually contain the information they seek.

Writing with new technologies has already become more than just plain text. Many scholars are encouraged to begin the advantage of the flexibility offered by new technology to include multimedia elements such as hypertext links, video and audio, and interactive elements in their electronic publications. And many students want the freedom to experiment with these new forms.

2.3 Responsibilities of the Project

- The move from paper to high quality electronic versions of theses and dissertations.
- Providing search capabilities can help to make information more readily available to scholars and researchers.
- The entire theses collections should be made available online.

3. Workflow and Process

3.1 Workflow

3.1.1 Metadata

Metadata is data about data. The term refers to any data used to aid the identification, description and location of networked electronic resources.

As ETD-MS is an interoperability metadata standard for electronic theses and dissertations, IISc has adopted. ETD-MS, which is an international standard for thesis and dissertation metadata developed by NDLTD (Networked Digital Library of Thesis and Dissertations). Dublin Core elements are the basis of ETD-MS although it provides specific to theses and Dissertations.

3.1.2 Metadata Elements of etd@iisc

Field Name	Value to be filled
creator	Author of the thesis
Contributor. advisors	Guide or advisor name
title	Title of the thesis
date. submitted	Submission date of the thesis
identifier.srno	SR Number of the student
subject.classification	Subject category of the thesis
subject.keyword	Keywords in the thesis
description.abstract	Abstract of the thesis
thesis.degree.name	Name of the Degree
thesis.degree.level	Level of the degree
thesis.degree.discipline	Discipline of the degree
description.note	Additional note about the thesis
thesis.degree.grantor	Grantor of thesis

3.1.3 Metadata elements pre-filled and hidden

Common fields are pre-filled and made hidden to save the time of the submitter.

Field Name	Value Filled
publisher	Indian Institute of Science
rights	Right statements about the thesis
type	Electronic Thesis and Dissertation
language	English
thesis.degree.grantor	Indian Institute of Science

3.1.4 Mandatory metadata fields

Some of the metadata fields are mandatory. System validates the mandatory fields at the time of submission.

Field Name	Value Filled
creator	Author of the thesis
contributor.advisor	Guide/Advisor of thesis
title	Title of the thesis
date.submitted	Submission date of the thesis
subject.keyword	Keyword in the thesis

3.1.5 Subject Classification

To enable the submitter to include the thesis under the most appropriate subject heading etd@iisc provides a classification scheme based on Dissertation Abstracts International (DAI).

3.2 Process involved in scanning of Thesis:

3.2.1 Scanning:

Scanning is done with the desired quality of 600 dpi and saved as TIFF format files and in the folder named OTIFF. The OTIFF folder size may be of 30-40MB.

Scanner : Minolta PS 7000
Software : Abby Fine reader 7.0
Manpower : One person may scan an average output of 2500-3000 pages per day

3.2.2 Quality Maintenance:

The OTIFF folder images are processed for print quality by removing fine noise, filling the unclear text, de-skewing, etc. The processed images are resized to the standard original text files and saved as PTIFF files.

This processing will also reduce the file size of OTIFF to 20-30 MB.

Software : Abby Fine reader 7.0
Scan-Fix 4.2.1

3.2.3 OCR and Links:

The processed images are converted as text by using OCR tool and for giving Links and searching text.

The final output will be saved as PDF file format as it is widely used format for electronic publishing in general. The file size of PDF file would be around 15-20MB.

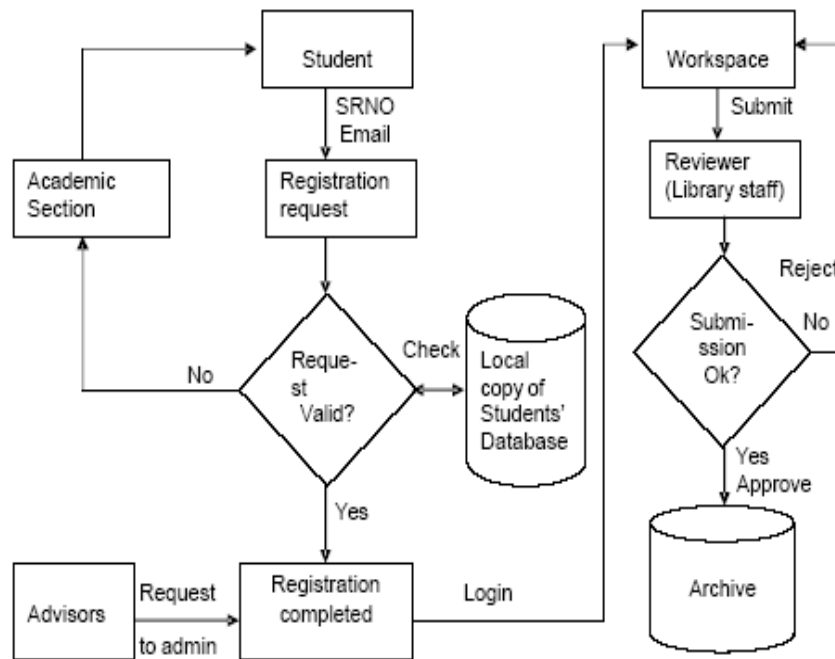
Software : Adobe Acrobat 6.0 Professional

4. DSpace- Digital Repository Software

DSpace (DSpace website, 2004) is an open source digital repository system that captures, stores, indexes, preserves, and redistributes an organization's research output in digital form. As institutional repository software DSpace is making its mark, with an increasing number of institutions around the globe installing, evaluating it for managing their digital assets. DSpace provides long-term physical storage and management of digital items in a secure, professionally managed repository.

DSpace is the first open source digital repository system to tackle the complex problem of how to accommodate the differing submission workflows needed for a multidisciplinary system. It has a strong and flexible administrative and security features like e-mail/password based authentication, e-mail notifications in different workflow steps and persistent identifier (handle) assignment for each item archived. DSpace provides both simple and advanced search and browse features. It supports full text search and thumbnail display of images in search results.

ETD Work Flow:



[Figure 1: etd@IISc workflow]

(figure taken from the article of reference no.6)

4.1 Online submission

In order to submit a thesis students have to register to etd@iisc. Using SR No. and valid email ID. SR No. and other information are validated against student’s database maintained by Academic Section. This database has been updated periodically once registration is done, a student can login using their email id and password to submit thesis.

4.2 Display of Communities and Collections Strengths

IISc has six divisions and each division has several departments. The divisions constitute the communities in etd@iisc and collections in each community reflect the departments of that division. Display of Communities and Collection strengths has been implemented.

4.3 Browsing features in etd@iisc

Browsing features have been provided for the following fields.

Author
Guides
By Date
BY Subject
Communities & Collection

5. Conclusion and Future works

The quality of an Academic Institution is reflected by the quality of its student's intellectual products. As the thesis literature was not available to the public, theses are little used and little cited. The thesis literature should be made available to the public so that the findings may be used in practical life. Digital image technology offers distinctive advantages to Institutions with impressive collection of scholarly resources. Primary long-term objective is to convert all these into digital format. Information content can be delivered directly to the reader without human intervention. Readers can retrieve information content in digital form remotely. Based on the value, use and legal issues Institute PhD Theses has been selected for digitization. The Archive of ETD enables the Students to submit theses electronically to the Archive. The archive has been developed to capture, disseminate and preserve research theses of Indian Institute of Science. Digitization is a method of preservation, and use Web Access as Publication. Paper documents can easily be produced from electronic documents but not vice versa. Digitization of existing Theses collection is already in progress. The ETD system generates a unique identifier (uri) to each thesis. Hyperlink to theses uri's has to be provided to our Web OPAC. It aims to evolve as a National Repository of Theses and Dissertation in collaboration with other Academic Institutions.

6. References

1. <http://etd.vt.edu>
2. <http://www.nedcc.org>
3. Shalini R Urs, TD Initiatives in India- Proposed Mysore University ETD project. Paper presented in *Workshop on an International Project of Electronic Dissemination of Theses and Dissertation held by UNESCO at Paris on 27-28 September 19*. Available at http://www.unesco.org/webworld/etd/september_99/contributions.html
4. Amit Kumar Bandyopadhyay, Digitization of Dissertation for Wide Accessibility: A Case Study of Faculty Opinions, P 376-383. Paper presented in the *4th International conference of Asian Digital Libraries at Bangalore on December 10-12 2001*.
5. Tony Gill, Anne Gilliland-Swetland, and Murtha Baca: Introduction to Metadata – Pathways to Digital Information.
6. P.Jobish, Hemanta Kumar Biswal, Filbert Minj, T.B. Rajashekar and S.Venkadesan – etd@IISc: A DSpace-based ETD-MS and OAI Compliant Theses Repository Service of Indian Institute of Science. Paper presented in the ETD2005: evolution through discovery, 8th International Symposium on Electronic Theses & Dissertations, 28-30 September 2005 at Sydney, Australia.

About Authors:

Mrs. K. Nirmala Devi is Assistant Librarian at Indian Institute of Science, Bangalore-560 012, INDIA and holds MLISc. Earlier she worked with NCSI, IISc. She is involved in Digitization activities of theses, journals etc., She is also involved in Digital Library of India activities. Her current interest includes setting up of ETD Archives and Digital Library.
Email:nirmala@library.iisc.ernet.in

Dr. S. Venkadesan is Librarian at Indian Institute of Science, Bangalore-560 012, INDIA. And holds B.Sc., B.E., M.S. and Ph.D. in Metallurgy from IIT. Madras. As a Metallurgist by profession, he was also worked as Head LIS at IGCAR, Kalpakkam and has undertaken several IT related projects and a Digital Library he is involved in providing access to a comprehensive print as well as online Digital information to the research community at IISc. He is also actively involved with the Digital Library of India activities. His current interest includes Distributed Archives, Digital Library, Internetworking and On-line resource sharing.
Email: venky@library.iisc.ernet.in

Filbert Minj is Scientific Officer, NCSI, Indian Institute of Science, Bangalore-560 012, INDIA. And holds M.Tech (Computer Science) from Jawaharlal Nehru University (JNU), New Delhi,
His areas of interest are Information Retrieval, Digital Libraries, DBMS
Email: filbert@ncsi.iisc.ernet.in