

# **etd@IISc: A DSpace-based ETD-MS and OAI Compliant Theses Repository Service of Indian Institute of Science**

**P. Jobish, Hemanta Kumar Biswal, Filbert Minj, T.B. Rajashekar**  
National Centre for Science Information, Indian Institute of Science, Bangalore,  
India

**S. Venkadesan**

JRD Tata Memorial Library, Indian Institute of Science, Bangalore, India

*Keywords:* institutional repository; digital asset management; OAI-PMH; electronic thesis and dissertation; community and collection strengths

## **ABSTRACT**

The Indian Institute of Science (IISc), Bangalore, was started in 1909 and is considered among the finest institutes of its kind in India today. ePrints@IISc, the open access repository of IISc research publications was established towards the end of 2002 using EPrints.org software. It has over 2000 research publications, is growing steadily and is accessed significantly from around the world today. Recently need was felt for setting up an ETD repository service for IISc. We decided to use DSpace repository software, given the recent positive experiences with this software for such services. We set up a prototype repository service to facilitate formulation of requirements for the production service and also address technical issues. Prototype helped us to obtain feedback from the IISc e-resources advisory committee and the concerned administrative section dealing with theses submissions and approvals. This led to identification of several key requirements for the production ETD service. These include: system will support only post-approval online submission of theses and not initial submission and refereeing processes; reflection of IISc divisions and departments as communities and collections; compliance to ETD-MS metadata standard; validation of student registration using student records database; automatic collection assignment; automatic metadata assignment and validation during online submission; support for assignment of subject categories; metadata and full text quality assessment by library staff; and issue of e-mail notifications to concerned parties during submission, approval and archiving processes. We successfully launched the etd@IISc service very recently by developing and implementing solutions to meet most of these requirements (<http://etd.iisc.ernet.in/>). We discuss the details in the paper. A unique customization we have implemented is display of community and collections strengths. etd@IISc is OAI-compliant and assigns unique handle-based identifiers. We also briefly discuss further work planned, including development of standard theses template for use by research students.

## **1. INTRODUCTION**

The Indian Institute of Science (IISc), Bangalore, was started in the year 1909 and is considered among the finest institutes of its kind in India today. IISc has more than thirty departments where active research is going on in the areas of science, technology and medicine. The outcome of the research includes theses of Ph.D students of various disciplines, research papers and dissertations of MSc. Engg. and M.E. students. ePrints@IISc, the open access repository of IISc research publications was established towards the end of 2002 using EPrints.org software. It has over 2000 research publications, is growing steadily and is accessed significantly from around the world today. Theses and dissertations represent the intellectual output of an academic institution. Indian Institute of Science has been producing valuable research outputs of high standard and significance from the date of its very existence. It is important that, reputed research institutes like IISc to publish and disseminate its research output through open access repositories. This will significantly enhance visibility of research work being carried out. Recently, a need was felt for setting up an ETD repository service for IISc. We decided to use Dspace repository software, given the recent positive experiences with this software for such services.

## **2. SOFTWARE SELECTION (DSPACE)**

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. It is adaptable to different community needs with different workflows, different levels of access control mechanisms and content moderation procedures. It has a comprehensive and customizable Dublin Core metadata collection for describing items individually. The authorization processes from the registration to submission and archival makes DSpace unique from other software.

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. DSpace is compliant to OAI-PMH protocol for metadata harvesting (Richard, 2004 and William, 2003).

## **3. PROTOTYPE REPOSITORY DEVELOPMENT**

Initially we set up a prototype repository service to facilitate formulation of requirements for the production service and also address technical issues. The aim behind the prototype development was to understand different needs and requirements of IISc research community towards the proposed ETD repository. The prototype development was focused to address the following questions:

- Examine recommended ETD workflow
- What policies are required for submission, approval, copyright and access management?

- Prepare draft policies for IISc.
- What document's formats are acceptable and why?
- Prepare draft document format policy
- Test for OAI compliance and interoperability
- Set up and configure ETD system supporting workflow, policies and document formats

Discussions and meetings were conducted with IISc e-resources advisory committee and the concerned administrative sections dealing with these submissions and approvals to evaluate the prototype repository service. This led to identification of several key requirements for the production ETD service. These include:

- System should support only post-approval online submission of theses and not initial submission and approval processes
- Reflection of IISc divisions and departments as communities and collections;
- Compliance to ETD-MS metadata standard
- Validation of student registration using student records database
- Automatic community and collection assignment to students on registration
- Automatic metadata assignment and validation during online submission
- Support for assignment of subject categories
- Metadata and full text quality assessment by library staff; and
- Issue of e-mail notifications to concerned parties during submission, approval and archiving processes

#### **4. IMPLEMENTATION OF PRODUCTION SERVICE**

Incorporating most of the requirements evolved from the prototype repository development, the theses repository of IISc, `etd@IISc`, has been implemented as a production service since February 2005. The current workflow of the system is as shown in the figure 1.

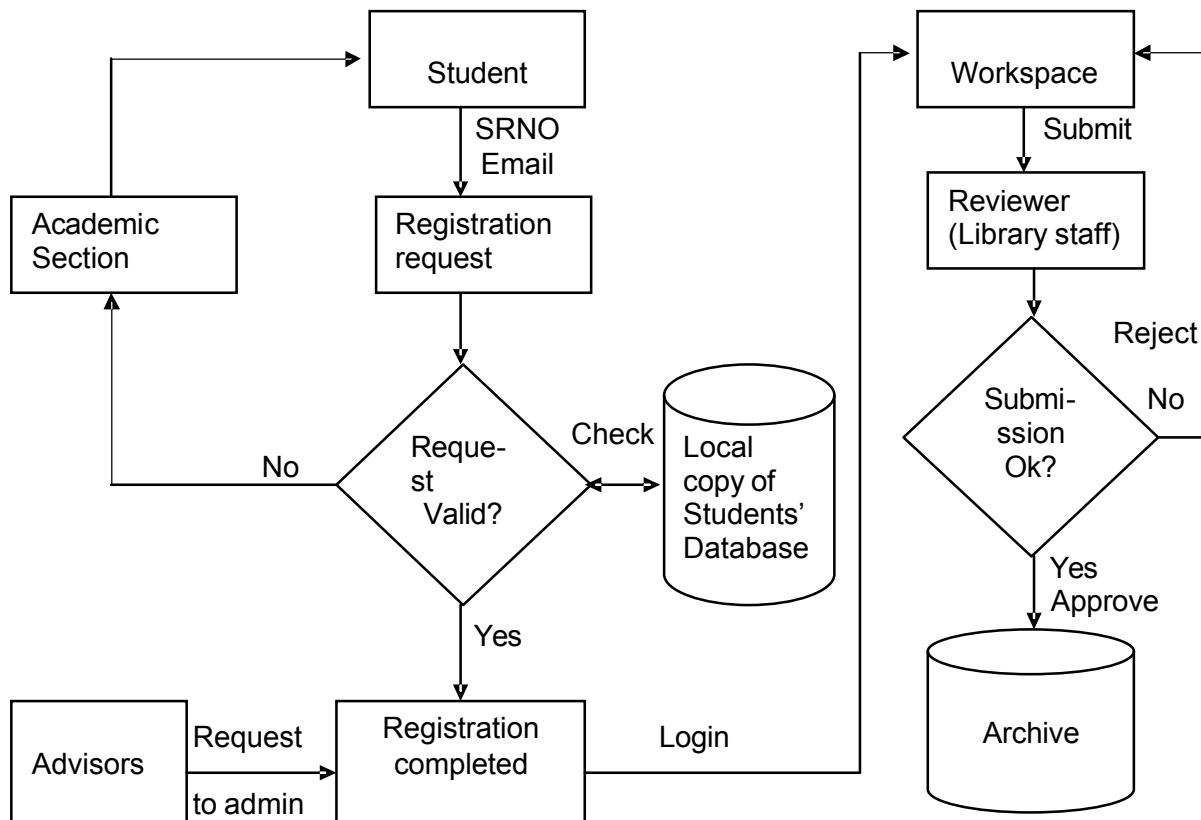
Additional functionalities added in `etd@IISc` to meet some of the above requirements are described below.

##### **4.1 Displaying 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. We have developed a patch to display community and collection strengths. The patch was incorporated in the code base of the release of Dspace (v-1.2.2).

##### **4.2 Registration / Submission**

Students need to supply Students Registration Number (SRNO) an unique student identification number and email to register with ETD system. It validates SRNO and other information against locally hosted students database, maintained by academic section. The locally hosted database gets updated periodically. Advisors (thesis guide) request admin to get registered into ETD system. It automatically assigns the collection to the registrant based on his registration details. The registration and the collection assignment are strictly based on the verification of student details in the database. Most of the submission forms are pre-filled during the submission process using content from the database. PDF is the preferred document format for thesis submission in `etd@IISc`.



[ Figure 1: etd@IISc workflow ]

### 4.3 Automatic E-mail Notifications

etd@IISc system sends e-mail notification to the submitter, workflow persons, and to the administrator at various stages of thesis submission. It also notifies the administrator, student, advisor and academic section about the archival of a thesis.

### 4.4 Subject Classification

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

### 4.5 Metadata Elements Used in etd@IISc

In addition to the default Dublin Core metadata registry, we have incorporated ETD-MS (NDLTD Web Site, 2004) metadata fields to make the theses repository ETD-MS compliant. This helps to give a richer set of metadata for the theses in the repository. The newly added metadata fields are listed in the table1. The submission GUIs and servlets codes were modified to incorporate the additional metadata elements.

Table 1: Newly added metadata fields

Field name	Value to be filled
thesis.degree.name	Name of the thesis
thesis.degree.level	Level of the degree
thesis.degree.discipline	Discipline of the degree
thesis.degree.grantor	Grantor of thesis

#### 4.6 Browse by Subject and Advisor

Apart from the default browse views, the need was felt to support browse by subject and advisor. We have implemented these browse features in etd@IISc.

#### 4.7 Handle Server Configuration

etd@IISc creates persistent identifier for every submitted thesis. This is unique, and can be accessible from any where around the world. The handle prefix for etd@IISc assigned by CNRI is '2005'.

#### 4.8 OAI Compliance and Interoperability

etd@IISc is compliant to OAI-PMH V-2.0 and ETD-MS V 1.01 to the metadata standard for theses and dissertations. It exposes the Dublin Core as well as the newly added metadata of items that are publicly accessible. OAI-Cat framework of OCLC is used to provide this functionality. It supports deletion information for withdrawn items and OAI-PMH resumption tokens. etd@IISc supports all the six OAI-PMH queries. Also the repository is registered with harvesting services like ARC, OAIster etc. to facilitate IISc research output to the international community.

#### 4.9 Help Manuals and Templates

Simple step-by-step help instructions have been provided for thesis submission, including instructions for creating PDF version of the thesis. The style file in MS word and Latex format is available on the repository site for thesis template. Also the repository provides links to some of the national and international ETD services for easier availability of theses worldwide from the desktops.

### 5. FEATURES OF ETD@IISC

We have made modifications and customizations in DSpace source code to facilitate the specific requirements of IISc. The features that we count unique for etd@IISc are:

- Reflection of IISc divisions and departments as communities and collections
- Compliance to ETD-MS metadata standard for theses and dissertations
- Unique registration process based on Student SR Number and IISc e-mail address
- Validation of student registration using student records database
- Automatic collection assignment of student during registration
- Automatic metadata assignment and validation during online submission
- Common fields are pre-filled and made hidden to save the time of the submitter
- Support for assignment of subject category
- Metadata and full text quality assessment by library staff

- Issue of e-mail notifications to concerned persons during submission, approval and archiving processes.
- Display of community, collection and repository strengths
- Browse views for advisor and subject fields
- Simple step-by-step help instructions for thesis submission, including instructions for creating PDF version of the thesis.

## 6. CONCLUSIONS AND FURTHER WORK

etd@IISc is a digital repository of theses and dissertations that runs parallel to the printed repository of theses and dissertations of IISc and facilitates better means to capture, store, process, and disseminate the intellectual output of IISc. Online access to IISc theses will give improved visibility for IISc research complementing the IISc e-print archive. etd@IISc has become operational very recently and we are observing the various operational implications of the repository and are very keen to incorporate further improvements.

We are exploring the possibilities of incorporating the following additional features for future works:

- Handling special characters during metadata submission
- Controlling access to PDF documents by out of IISc on selective basis (if required)
- Provision for duplicate checking or duplicate alert mechanism
- Supporting deferred access to full text thesis based on student/advisor request - e.g. initial 6 months or 1 year.
- Development of standard theses template for use by research students

## 8. ACKNOWLEDGEMENTS

We acknowledge Aneesh Joy for supporting the development of browse views by subject and advisor.

## 8. REFERENCES

1. Jobish, P & Biswal, Hemanta Kumar. *Implementing IISc E-Theses system using DSpace and ETD-MS Metadata standard for theses and dissertations*. Unpublished project report, NCSI, Indian Institute of Science, Bangalore, India.
2. Richard Jones.(2004, Jan 30). DSpace vs. ETD-dB: Choosing software to manage electronic theses and dissertations. *Ariadne*, (38). Retrieved Aug. 20, 2004 from <http://www.ariadne.ac.uk/issue38/jones/>
3. Smith, MacKenzie, et al. (Jan, 2003). DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, 9(1). Retrieved Oct. 3, 2004 from <http://www.dlib.org/dlib/january03/smith/01smith.html>
4. William Nixon. (2003, Oct 30). DAEDALUS: Initial experiments with EPrints and DSpace at the University of Glasgow. *Ariadne*, (37). Retrieved Aug. 30, 2004 from <http://www.ariadne.ac.uk/issue37/nixon/>
5. *Dspace website*. Retrieved August 20, 2004 from <http://www.dspace.org>
6. NDLTD Web Site. Retrieved August 20, 2004 from <http://www.ndltd.org/standards/metadata/current.html>