

# Trends in the development of E-Theses in India: issues, constraints, and solutions

K V Satyanarayana\*, B Ramesh Babu\*\*

\*[k.satyanarayana@tcs.com](mailto:k.satyanarayana@tcs.com), Tata Research Development & Design Centre, Tata Consultancy Services Ltd. (Pune), \*\* [beeraka\\_r@yahoo.co.uk](mailto:beeraka_r@yahoo.co.uk), Department of Information Science, University of Madras (Chennai)

## Abstract

This paper summarizes current developments of e-theses in India and discusses barriers to implementation and proposes a model plan to speed up the process. If recent guidelines from the University Grants Commission of India encouraging the submission of e-theses for deposit are universally adopted; India will see a proliferation of institutional repositories in the country. There are a few institutes such as the Indian Institute of Science, Indian Institute of Technology, Mumbai (IITB) already spearheading the deposit process whereas a few others like Indian Institute of Technology, Delhi, National Chemical Laboratories, University of Hyderabad and some other public and private institutions have joined the deposit process. Current national level initiatives such as bibliographic records of PhD theses at INFLIBNET or the Vidyanidhi project at University of Mysore, designed to act as a national repository for e-theses are yet to become comprehensive and authentic sources of information for all PhD theses submitted to Indian Universities. Also, e-theses repositories pose major challenges to Libraries and to research scholars. Some of these challenges include concerns about plagiarism, copyright violations, quality of theses, lack of policy changes at university level, poor infrastructure, inadequate technical skills of library staff in installation, maintenance, and customization of institutional repository software, knowledge of UNIX and Linux operating systems, and limited understanding of using metadata schemes and copyright issues. Current institutional repositories training programmes in the country are inadequate in number and only expose the participants to such software and do not empower the participants with the practical knowledge of using these software. Another significant issue is the lack of adequate and suitable programmes available to train working library professionals in IPR issues. An action plan is proposed to spearhead e-theses in India which includes short-term training on open source software and IPR for library staff of university and R&D institutes, availability of customized software with innovative features as seen in sites like ETD@IISc, creation of paid or free consultation facility and blogs, creation of templates for Web sites for thesis submission, recommended classification scheme for thesis to explore systems like Urkund or Genuine Text for locating potential plagiarism. It is further suggested that institutes like NCSI, DRTC, INFLIBNET, and a few others engaged in training library professionals can play a significant role in e-theses depositories growth in India.

## 1. Introduction

Electronic Thesis and Dissertation (ETD) initiatives in India are at an embryonic stage. At present, access to theses and dissertations is limited, as availability of comprehensive theses bibliography database or ease of availability of full-text theses is not available. Since the beginning of this decade, few national institutes started ETD initiatives with focus on Open Archives Initiative (OAI) compliant repositories, thus allowing metadata harvesting. It may not be surprising that some of these institutes received funding or collaborative support from other organizations to start their digital library initiatives. The Central University of Hyderabad has collaborated with Sun Microsystems Ltd. to digitize its collection and has also received the hardware at a discounted price. IIT-Kharagpur has been awarded a project by the Aeronautics Research and Development Board for developing a digital library on aerospace and technology. ETD@IIT Delhi has received grants from the Ministry of Human Resources Development and the Department of Biotechnology for a similar initiative. On the other hand, the Information Library Network (INFLIBNET) consortium, Vidyanidhi Library of Theses and Dissertations etc. are actively involved in the development of bibliographic ETDs and digital ETDs.

Last year, UGC pronounced some guidelines towards the submission of Metadata and Full-text of doctoral theses in the Electronic Format, to improve creation, collection and

compilation of cataloguing information, and to create the Indian National Theses Database. This will likely to drive new ETD initiatives in India.

## 2. Current Scenario

Currently in India, there are less than 20 digital ETDs available. The Indian Institute of Technology-Bombay has introduced online submission of electronic theses and dissertations of postgraduate and PhD students since January 1999. This may be the first such attempt in India. The following year the Vidyanidhi project started.

ETD initiatives in India can be broadly grouped under the following three categories :

1. Institutes creating ETDs by using open source software
2. Institutes creating ETDs by using third party/in-house software or web technologies
3. Aggregators trying to create ETD bibliographies or full-text ETD collection

Table 1: Open source software ETDs

Institute Name	No. of ETD documents	Software Used
Indian Institute of Astrophysics	83+	DSapce
Indian Institute of Science, Bangalore	227+	DSpace
Indian Institute of Technology, Delhi	30+	DSpace
Indian Institute of Technology, Kanpur	9300+	DSpace
Indian Institute of Technology, Kharagpur	NA	DSpace
Indian Institute of Technology, Mumbai	3500+	Greenstone
Indian Statistical Institute, Bangalore	NA	DSpace
Librarian's Digital Library, Bangalore	3+	DSpace
National Chemical Laboratories, Pune	208+	DSpace
National Institute of Oceanography	20+	DSpace
National Institute of Technology, Rourkela	13+	DSpace
University of Hyderabad	4000+	DSpace

Interestingly, it is observed from Table 1 that DSpace is the most popular software for ETDs in India. It was reported that IIT-B is also in the process of migrating its ETD from Greenstone to DSpace.

Table 2: ETDs with in-house software or no software

Institute Name	No. of ETD documents	Software Used
National Institute of Technology Karnataka, Surathkal	NA	NA
National Institute of Technology, Calicut	1000+	NA
Physical Research Laboratory, Ahmedabad	75+	HTML display
SRM University, Chennai	450+	In-house

At least one institute from the list in Table 2 reported that they would be moving to open source digital repository software.

Table 3: ETD Aggregators

Institute Name	No. of ETD documents
INFLIBNET	153206+ bibliographic documents
Vidyanidhi	50,000+ bibliographic records and 4151+ full text theses
DELNET	44,304 records+ bibliographic records

Though, these aggregators try to provide bibliographic access to a larger extent and to full-text theses to some extent, they have failed to meet the expectations of the research community and to provide single interface to Indian Thesis. The Association of Indian Universities publishes a list of doctoral thesis regularly.

Apart from these, there are a few other digital repositories which mention about ETDs but have either one or no theses in their collection. Such repositories are from the Indian Institute of Management - Kozhikode, Raman Research Institute, National Aeronautics Laboratory etc.

Some of these ETDs provide notable customizations and access improvement mechanisms. IIT-K embedded its logo in each page of thesis, customized browsing option by supervisors and citations and provided cross references to IITK theses. IISc partnered with free science-specific search engine, SCIRUS, to index its two institutional repositories, ePrints@IISc and etd@IISc. It also offered a selected list of publishers' policies for local use. Author benefits and copyright issues were also detailed on the website. Display of individual collection strengths was also developed at NCSI. The Dissertation Abstracts International classification scheme was used. IIT-Kanpur claimed that once it put-up on the web, it would be the largest OAI compliant ETD repository and one of the top ten in the world.

### 3. Issues

While there is a progressive growth of digital libraries in US, Australia and many countries in Europe, India is witnessing moderate growth. Indeed, many ETD initiatives are from institutes of national importance. Universities and other academic institutes are lagging behind in this direction.

#### 3.1 High volume of research output

There are over 350 universities and research institutes and more than 15,000 colleges in India. Over 60,000 researchers are enrolled in a year. However, statistics regarding doctoral research output is difficult to get as there is no single mechanism of cataloguing the data despite the Association of Indian Universities publishing these information in its newsletter. There are estimates ranging from 11,500 theses per year to 25 to 30,000 doctoral theses produced in India. The volume of theses output per year will be much more if we consider masters theses. If we need to consider, retrospective conversion of all masters and PhD theses, the number can be estimated to be in the order of 20-25 times than the annual output. Also, this problem is further complicated by multilingual content of theses as digitization of these will provide different challenges.

#### 3.2 IPR issues

Global online access of thesis content in full-text form is still an issue among the academic community. We had come across an example where one university despite having thesis in the electronic format did not create a digital repository due to copyright issues. Some of the existing ETDs provide access to their theses in the intranet only and in few cases, provide hardcopy access only. Indeed, many ETDs do not provide guidance to researchers regarding journal and book publisher's policies concerning pre-posting of digital dissertations. Also, it is not clear whether the embargo period policy is available and practiced by these ETDs. Having such an embargo policy is critical from four perspectives – potential legal issues for institutes arising out of providing access to copyrighted information in thesis, potential patent disclosure and publication issues for authors with online posting of theses and for corporates, its inability to share results of sponsored research projects with others immediately.

#### 3.3 Plagiarism

Few surveys on using digital ETDs raised concerns regarding plagiarism and misuse. A study conducted at Kerala Agriculture University and at Indian Institute of Spices Research found that over 89% of scientists and 81% of students think that online access to full-text theses increases the chances of misuse of documents. At the same time, 56% of scientists and 43% of students suggested that a 2-3 year embargo period would avoid any misuse and would allow time for researchers to document the thesis properly and publish research findings.

#### 3.4 Quality of research

It is generally believed that the quality of doctoral theses produced in Indian institutions varies widely and not all theses result in significant research publications in journals of high relevance. In the same Kerala Agriculture University survey mentioned above, 80% of scientists and 78% of students felt that wider access to theses will improve wider discussion on the research findings, thus, indirectly improve the quality of further research and reports and theses.

### 3.5 Preservation

A review of current literature on Indian ETD case studies does not reveal much about their archival and long-term preservation policy. Long-term preservation is a concern for any digitized material and ETDs are no exception. IIT-B has mentioned clearly in their archiving policy that it will archive the databases through more than one back-up method. It maintains CD-ROM backup, DAT Tape cartridge backup, and ETD mirror server maintained at different location. Long-term commitment is required in building and managing repositories.

### 3.6 Limited support from Librarians

J K Vijayakumar, TAV Murthy and MTM Khan surveyed 65 university librarians, 171 researchers and 77 research guides and presented their results at the International CALIBER 2005 in Feb 2005. According to them, 83% of researchers are supportive of online access to PhD theses, whereas 48% (31) of librarians are supportive of such ETD initiatives. These authors believe that gaps in national policies and frameworks may be reasons for librarians' non-supportiveness. With the new UGC regulations, librarians should be more willing to initiate ETD initiatives in future. The good news is that 34% (22 universities) of surveyed universities have already adopted a policy to collect e-format of theses along with the submission of printed copies. 18 university libraries (28%) already started creation of ETD systems.

### 3.7 Lack of trained staff

It may not be surprising to see that university librarians' positions have been vacant for many universities. Our interaction with a few librarians indicated the lack of technical awareness of digital repositories. Some of them lack awareness of digital software used in their organization.

### 3.8 Limited number of Training Programs

There are not many training programs available for learning installation and maintenance of digital repositories. Most of the library professionals may lack the technical expertise for customizing the software. While organizations like NCSI, DRTC, DELNET, INFLIBNET, IIM-K may organize these programs. These are not sufficient in frequency and are also organized in very few cities, where the university Library staff has to travel long distances to attend such programs. We had come across a couple of training programs where a majority of participants were from non-academic institutes. Currently library professionals depend on NCSI/DRTC subscription forums to seek answers to their technical problems. DRTC has prepared a shell script for the installation of DSpace. However, technical support at a sustained level to all Institutes is a big concern today.

### 3.9 Limited success of aggregators

INFLIBNET was able to provide searchable free databases PhD theses but it is yet to be far from being comprehensive and single source of theses database. Similarly Vidyanidhi, in existence since 2000, made limited success in attracting full text theses and was unable to gain a status of a national digital theses library database on the lines of UGC's proposed National Digital Library.

Apart from these issues, open access to full-text theses is uneven in Indian ETD repositories. Some of these repositories are not available on the intranet. Also, they do not provide access statistics to encourage the use of ETD repositories and to show increase in the usage thesis.

While bigger institutes may not be facing problems in terms of funds, smaller institutes complained of budget issues. In 2005, P. Vinayagamoorthy, B. Rameshbabu and S. Gopalakrishnan analyzed 225 institutes from Tamilnadu for their Digital Library initiatives in Engineering Educational Institutions. This is one of the largest sample surveys conducted in India in understanding the digital library initiatives. They identified a few reasons for hindrances to non-development of digital repositories. They are initial cost (77%), insufficient funds (50%), IPR issues (57%), security issues (44%), lack of adequate trained staff (38%) and preservation (39%). These reasons will be definitely valid across the country and across all institutes.

## 4. Other International Initiatives

Let us look at few national programs in countries such as in Australia, Canada, and EThOS in UK to understand their working models.

### 4.1 Australian Digital Theses Programme (ADT)

This is a collaborative initiative and universities voluntarily participate in the program. The initial project was funded by the Australian Research Council (ARC). The original seven partners continue to guide and advise the national group in their role as the ADT Steering Committee. ADT provides free software to members. Members deposit or submit digitized theses onto their local servers. Metadata is harvested onto a central server providing searchable database. Full-text theses files reside on the local institutions' servers.

### 4.2 EThOS Project

This project is sponsored and funded by The Joint Information Systems Committee (JISC) and Consortium of Research Libraries in the British Isles (CURL) and is based at the British Library. This model brings a fundamental shift from the current situation where the end-user/institution pays for the thesis request, to one where the cost is borne by the institute that supplies the thesis. Sponsors institutes pay towards the annual subscription and these funds are utilized for digitization of their theses collection. Another category called Associate Institutes pays for digitization of theses in response to requests from end-users. While users are provided electronic copies free of cost, alternative formats are charged. All sponsor institutes are involved in various aspects of the EThOs project. EThOS makes available a metadata of over 500,000 records from the earlier systems.

### 4.3 Theses Canada

A centrally coordinated program for theses started in 1995 as part of Library and Archives Canada. To date it continues to be a voluntary program. It has more than 250,000 masters' theses and doctoral dissertations in its collection. Thesis Canada harvests metadata and ETDs if the university have electronic submission programs, otherwise, universities can send physical theses to ProQuest Information and Learning for processing. Currently Theses Canada pays approximately one-third of subsidy process charges.

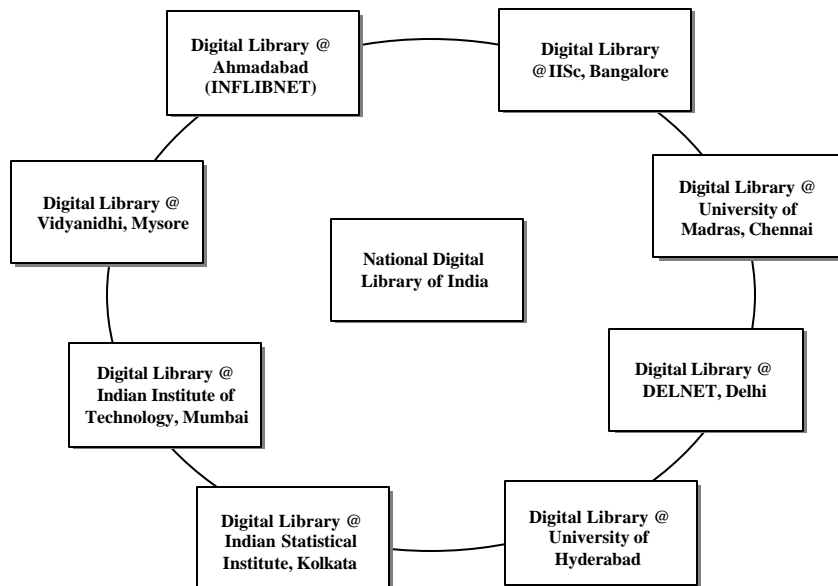
## 5. Action Plan

Current initiatives in India are unable to provide any highly successful stories in India. There may be several possible models that can be considered, but given the high volume of masters and doctors theses output every year in India and huge backlog of retrospective conversion of theses, we would like to propose that a virtual National Digital Library of India (NDLI) with multi-site locations in Delhi, Kolkata, Mumbai, Chennai, Bangalore, Hyderabad, Mysore and Ahmedabad. NDLI will be governed by a Project Board consisting of distinguished personnel from the fields like library science, education, technology and law. They will oversee the progress of the project and will help in getting partial funding from agencies like the University Grants Commission (UGC), All India Council for Technical Education (AICTE), Ministry of Human Resource Development and the Ministry of Science and Technology etc. which are responsible for funding higher education in India. This model can easily incorporate into UGC's planned Indian National Theses database

Its immediate mandate will be providing single interface to theses data from India and to start digitization of theses in English with high priority. Participation to the program will be voluntary. These centers will be able to provide digitization services either through in-house facilities or selected third party providers like millennium library project scanning centers like the one available in IIT-A. It is encouraged to institute an annual subscription for bigger institutes where thesis output is more than 10,000 till date. These institutes can opt for digitization services of the National Digital Library network providers. Other institutes can chose to opt for digitization services based on demand. NDLI will collaborate to get access to already available theses collection with organizations like National Social Science Documentation Centre (NASSDOC) which already maintain a huge collection of Social Sciences theses.

Possible candidates for these sites can be DELNET (Delhi), NCSI at IISC (Bangalore), Vidyanidhi (Mysore), INFLIBNET (Ahmedabad), Indian Statistical Institute (Kolkata), University of Hyderabad (Hyderabad), University of Madras (Chennai), Indian Institute of Technology (Mumbai).

Figure 1: Proposed National Digital Library of India model



One or two centers of NDLI will drive the following initiatives and other centers will be closely involved in these initiatives.

### 5.1 Marketing

Mandatory electronic submission of theses requires all universities to change their guidelines and this is a time-consuming exercise. Few recent surveys in India showed that researchers are willing to provide e-theses; all institutes should encourage voluntary submission of e-theses by researchers. This may require active marketing efforts by library with various departments in the institute. NDLI will provide the required marketing material to participant libraries and also manage the web-site for active promotion. NDLI will also provide the tools and templates required for thesis writing to researchers.

### 5.2 Training

To bridge the gap, NDLI will either conduct training programs across India, one program every month for next two years through its resources or will arrange for training programs through other institutes like DRTC, IIM-K etc.

### 5.3 Digitization standards

NDLI will define standards for digitization, will provide recommended hardware and software configuration, will select providers for outsourced digitization activities and will define SLAs for digitization of theses both for on-demand requests from researchers and for participant institutes.

### 5.4 Open Source Center of Excellence

NDLI will set-up *center of excellence* of open source digital repository preferably at National Centre for Science Information at Indian Institute of Sciences, Bangalore. NCSI has readily available technical talent pool and also involved in training programs. This center of excellence should have the following mandate:

1. **Standards:** It will suggest metadata recommendations
2. **Templates and Plug-ins:** It will develop plug-ins for different repositories for easy metadata harvesting in the required scheme; provide facility for RSS feeds, customized theses submission form templates, metadata entry mechanism for subjects like Chemistry where special characters are included. It will also develop templates for full document suppression or partial document suppression based on author recommendation and automatic availability of document after embargo period, develop batch processing methods and work on long preservation techniques and recommended migration methods whenever version changes of software happens.
3. **Unicode support:** This center will work on developing repositories in Indian languages with Unicode plug-ins for all major Indian languages.
4. **Expertise building:** Though at present, DSpace is largely used in India, it is encouraged to build capability in three major repository softwares, namely, DSpace,



Eprints and Fedora and also look into other repository software options. It will also work with other international initiatives to formulate preservation methods to be adopted in India. It will also handle dedicated help lines for handling user problems and launch a blog for open source repositories.

5. **Plagiarism and security issues:** This center will experiment with tools like Urkund, Genuine Text, Turnitin and iThenticate to find efficient ways to detect plagiarism. It will also work towards improving document security for implementing embargo policies and unauthorized access.
6. **Alliance building:** It will explore ways to seek help from IT companies where open source research groups or DSpace installations exist like Tata Consultancy Services and Infosys.
7. It will also harvest data from all repositories.

### 5.5 IPR issues

It is necessary for institutes to bring awareness in its researchers about existing copyright laws and implications of third party material available in theses. They need to encourage authors to seek permissions from any third-party copyright owners and inform authors of the need to edit their thesis before submission to the repository in those cases where copyright permissions is not available. Hence, one or two Indian Institute of Technology libraries should own the responsibility of working closely with the National Law University in India to create an IPR portal for researchers which can be accessible by all researchers in India. This site should also be able to provide publishers and patent office policies of preposting of theses.

### 5.6 E-commerce

NLDI will provide thesis in print or CD format as a value added service to researchers for a price. It will also tie-up with players like Google for providing access to Indian thesis database freely. In return, it can share the advertisement revenues generated out of page views.

Apart from these areas, other action points include - current Indian aggregators should find ways to avoid duplication of efforts in creating bibliography of theses, all ETDs to provide global access to bibliographic data and big ETDs to tie-up with SCIRUS, Google etc. to allow access to bibliographic data searching.

## 6. Conclusion

It is necessary for India to have a network of National Digital Library centers to spearhead the process of ETD movement. Further research is required to finalize the model and estimate the initial investments required.

## 7. References

Francis, A. T; Sushama Devi, C.K; and Abdul Razak, C (2007). Content management for digital delivery of agricultural information: redefining need of libraries in the context of digitization of theses and research reports. *In International Conference on Semantic Web and Digital Libraries (ICSD-2007)*. ARD Prasad & Devika P. Madalli (Eds.): ICSD-2007. [https://drtc.isibang.ac.in/bitstream/1849/370/1/AT\\_francis.pdf](https://drtc.isibang.ac.in/bitstream/1849/370/1/AT_francis.pdf)

Ghosh, S. B. and Das, Anup Kumar (2006). Open access and institutional repositories – a developing country perspective: a case study of India. *In Proceedings IFLA Council and General Conference (72nd : 2006 : Seoul), Korea*. <http://eprints.rclis.org/archive/00006391/>

Vijayakumar, J K; Murthy, TAV; and Khan, MTM. Indian academia on copyright and IPR issues of electronic theses and dissertations. *In International CALIBER 2005, Kochi, 2-4 February 2005*. Ed. Murthy, TAV and others. <http://eprints.rclis.org/archive/00005657/>.

Vijayakumar, J. K. and Murthy, T. A. V. and Khan, M. T. M. (2005). Introducing electronic theses and dissertations in universities: an Indian perspective. *In Proceedings GL7 : Seventh International Conference on Grey Literature, Nancy (France)*. <http://eprints.rclis.org/archive/00005670/>

Vinayagamoorthy, P, Ramesh Babu, B and Gopalakrishnan, S (2006). Digital Library Initiatives in Engineering Educational Institutions in Tamil Nadu. *Indian Journal of Technjical Education*, 29(1): 68-88

Wolf, Martin (2006). Electronic Theses Online Service (ETHOS) – A National UK Project. *In 9th International Symposium on Electronic Theses and Dissertations. IXe Symposium international sur les thèses et mémoires électroniques. June 7 – 10 Juin, 2006, Quebec City / Québec, Canada*