
ELECTRONIC THESES AND DISSERTATIONS: AN INDISPENSABLE TOOL FOR ACCESS TO GLOBAL RESEARCH OUTPUT

Atanu Kumar Sinha

Abstract

Providing access to information has always been a function of Libraries and Information Centres (L & ICs). Access to information is crucial for education, science and culture and for fostering democracy in the information society. The principle of 'free and universal access to information as well as freedom in the creation and dissemination of knowledge' is a fundamental goal of the L & ICs. Theses and Dissertations (TDs) among all the primary sources are probably least used because firstly, their existence is not known in many cases, and, secondly, they are not easily accessible due to their unpublished status. This paper focuses on access and technology issues concerning ETDs, nature and benefits, methods of developing digital archives, and discusses various file formats and file compression as well as post processing for making digitised files accessible to the end-users.

Keywords: Library Automation; Digital Library; ETDs; Digital Archives; Digital Access

1. Introduction

We are living in an e-world, a world permeated by the effects and products of electronic technology. E-business, e-learning, e-Government; 'e' is pervasive. It is interesting that how new terms have emerged in combination with the 'e' preceding to it. It is reasonable to assume that eventually e-activities will become redundant. In the e-world, e-encounters are commonplace. To fulfill the user's demands, the advanced libraries provide not only e-environment, e-resources and e-services but also fulfill user's role to use e-resources in the digital environment (Sinha, 2006).

Theses and Dissertations (TDs) are regarded as the bedrock of graduate education. As secondary source of information, TDs are known to be the rich and unique source of information. Often the only source for research work that does not find its way into various publication channels. The vast majority of these works are kept un-utilised in college and university libraries. The best way to bring this original research to the light is to publish electronically and give free and open access to these documents via the Internet.

In recent years, access to abstracts, journal articles, conference proceedings and book chapters in electronic format have become commonplace. As researchers have come to expect quick, easy access to information from any time, increasing demands have been placed upon resource providers. Expectations relating to the production and use of theses and dissertations have proved no exception to

the rule. At international and also in India a significant number of Universities and L & ICs have made doctoral and master level theses available in digital format to fulfill the principle of 'free and universal access to information' as well as freedom in the creation, treatment and dissemination of knowledge. 'Free exchange of ideas and knowledge' should be the main goal to redefine the universal access to information and maximum level of services to be provided to information users (Singh, 2005).

2. Internet and Library

In the Digital Networking environment, the Internet plays an important role in making the 'citizen' to 'netizen'. Libraries have always been considered as the channels of information communication. The value of the library is found in its collection of information resources. The concept of 'Library Automation' has emerged as a leading edge technological solution to the persistent problems of enhancing access, enduring archiving and expanding the dissemination of information. An automated library with the convergence of Internet offer a powerful means of managing information resources. The concept of "Library without Walls" has almost become a reality because of the Internet that provides a wide variety of digital information through web browser on user's desktop.

3. Theses and Dissertations

Theses and Dissertations (TDs) are by their very nature meant to report some original work in a specified field. One important characteristic of such research reports is that they may be the result of purely academic pursuit. Among all the primary sources, theses are probably least used because

1. firstly, their existence is not known in many cases, and,
2. secondly, they are not easily accessible due to their unpublished status.

One tends to consider theses only as primary literature. The extensive survey part and the bibliography can serve as a secondary source in searching for literature. A copy of the theses is usually kept in the library of the university granting the degree and subject to limiting the conditions, which are consulted/borrowed/photocopied. To find out what theses exist, one is needed to consult indexes/catalogues. In this context it may be said if these are digitised and may be obtained in the networking it will be easy for the researcher not to come to library. They may obtain the needed information without losing any time. Thus, an obvious challenge today is the problem of how to integrate electronic resources with more traditional forms. The need for complete integration seems taken for granted by L & ICs and the users alike at least in the scholarly community: as Dougherty and Hughes (1991) report: "Provokes the librarians... prefer a future in which there is universal access by faculty and students to multiple information sources in all possible media via a single multifunctional workstation" (Dougherty, 1991).

4. ETDs- Definition

Electronic Theses and Dissertations (ETD) is just like printed theses and dissertations that follows current e-paper format. An ETD is a theses and dissertation in electronic format that is archived and organised into a Digital Library (DL) collection and made accessible electronically via networks such as Internet. ETDs may be a born digital one or digitised version of born analog (print) theses. Maintaining a copy of the electronic version of theses does not automatically result in an archive of Digital Library of e-theses. A Digital Library is a collection of digital content and builds organisation

and services around the digital collection. A DL of ETDs is an infrastructure and a facility for creating, archiving, storing, accessing and disseminating ETDs.

5. Nature of ETDs

There are two standards of ETDs. These are-

1. author-created documents consisting of Word file converted to PDF, HTML or XML, and submitted (typically) over a network connection with related metadata; and;
2. electronic files created by scanning the pages of paper theses and dissertations.

The latter is usually created by university or information centres and the resulting documents are much less desirable than author-created documents because, they cannot be manipulated. Scanned ETDs also require more storage space.

6. Why ETDs

The process of scrutiny, validation and approval of ETDs is confined to few experts (identified by the university recommendation of theses supervisors). It is not opened to the scientific community at large, these are also kept in closed access, making these difficult for other students to access them. These remain un-tapped and under-utilised asset, leading to unnecessary duplication and repetition. Thus, ETDs are needed for the following significant advantages for students, university faculty and staff, and institutions as a whole.

6.1 Benefits for Students

- A student is likely to improve his/her IT skills in the process of creating an ETD.
- The opportunity to create a born digital theses allows a student to express his/her research results in creative and flexible ways.
- The production of an ETD is a cost effective alternative to printing and binding a large number of paper copies.
- Students' theses are likely to be read more widely if they are easily accessible via the Web.

6.2 Benefits for Academic Staff/Faculty:

- Researchers can undertake full content searches of ETDs on the Web, without having to judge from an abstract whether it is worth requesting a copy of the full text of a hard-copy theses on inter-library loan.
- Researchers are able to access ETDs on the Web as soon as they are required.
- ETDs can be obtained from remote locations at any time.
- a researcher can obtain access to an ETD regardless of the number of other academics using the same material at the same time.
- Researchers can use the possibilities offered by ETDs to motivate students and encourage them to think of the diverse ways in which their research results can be expressed.

6.3 Benefits for Institutions

- Access to the research output of staff and students is improved through the medium of ETDs. An increase in the level of use of such high level material is of benefit to the institution in terms of promoting its research profile.

6.4 Benefits for Libraries

- The benefits of ETDs from the library's point of view are-
- Better stocked Digital Library
- More timely access to current research
- Available ALL DAY/ 24 Hours
- Never checked out never overdue- no fines
- Serve more users with reduced staff
- Fewer physical copies to handle
- Less shelf space required for storage
- Less expense to ship them for microfilming
- Reduce processing costs
- Cataloguing record derived largely (programmatically) from existing e-text
- Nothing to send to the Binder
- Staff not needed to circulate and (re)shelve (RGUL, 2006; Singh, 2005).

7. Digital Library Vision

Digital Library may be considered as a versatile digital information repository. The structure of Digital Library falls into two categories:

- a. Preservation of Born Literature, and,
- b. Preservation of Made/Prepared Digital Literature.

The Digital Library vision is for single, continuously available, interoperable platform. The Digital Library visions are access to information via –

- Physical Access
- Intellectual Access
- Long-term (preservation) Access.

There are two parts of a Digital Library system-

1. Offline part - preparing a document collection for presentation, and,
2. Online part - presenting the collection to the user through an appropriate interface.

8. File Formats

There are various file formats. Different factors affect the choice of the format at each stage of the digitising process, such are-

Acquisition: This is the first and most important step as it maintains the highest fidelity to the original.

Archival Storage: There must be a standard format that will be readable in the future. The ability to hold associated metadata may be useful.

Editing: Proprietary formats may be useful to support any editing of information.

Delivery: Factors to be taken into account include destination device (screen, printer) and its capabilities, delivery method, file size and network bandwidth and format support at destination (Zhang, 1999).

8.1 General File Formats

There are two main general file formats. These are-

1. PDF (Portable Document Format) is a common file format used in many digital documents as with the PDF format one can get all the media, colours, graphics, fonts and formats. Adobe Acrobat Reader software is needed to read a PDF file.
2. HTML is another file format used in the Web and Digital Libraries. An HTML file may comprise text as well as non-text materials, such as, graphics, images and multimedia information. HTML file can be read by any web browser and also by software like MSWord.

8.2 Image File Formats

There are the following popular image file formats.

1. GIF (Graphic Interchange Format), developed by CompuServe is a device-independent method of storing pictures. It can store 8 bits/pixel (256 or fewer colours).
2. JPEG (Joint Photographic Experts Group) format stores full-colour information in 24 bits/pixel (16 million colours). It is commonly used for encoding digitised photographs.
3. TIFF (Tagged Image File Format) is a multi- purpose raster file format developed by Aldus and Microsoft with the objective of providing a basis for importing scanned images into the desktop publishing packages.

Table :1, Different File Formats in Digital Library

Type of Text	Abbreviation	Format	File Extension
Unstructured Text	ASCII	American Standard Code for Information Interchange	.txt
Structured Text	HTML XML PDF	Hypertext Markup Language Extended Markup Language Portable Document Format	.html .xml .pdf
Images	PDF BMP TIFF	Portable Document Format Bit Map Page Tagged Image File Format	.pdf .bmp .tif
Web Compatible Image File Format	GIF JPEG	Graphics Interchange Format Joint Photographic Experts Group	.gif .jpg

The preservation of library collection through IT is an easy task which follows following the two processes:

8.2.1 Scanning documents and making image file

In this process the scanning of documents is an initial process. The documents should be scanned in the resolution of 30 dpi (dot per inch) depending on the physical condition of the documents. The scanned images are typically stored in the file formats like JPIG (Joint Photographic Experts Group), TIFF (Tagged Image File Format). The final output is converted to a light format like PDF (Portable Document Format).

8.2.2 Scanning and preparing full textfiles

The scanned and processed images before being converted to PDF files, under the process to separate the text and media from the image files. The next process which is an important task is to rectify the errors and make text files error free with the help of OCR (Optical Character Recognition) process. This process also involves proof reading and checking quality of text files. After completion of all the above process, the text is converted into the mark up languages like XML, HTML, etc. along with the complete tagging which enables the databases to be integrated with any Web applications (Prasad, 2006).

9. The Process of Digitisation

Steps of digitizing process can be presented by a diagram as shown in the following section-

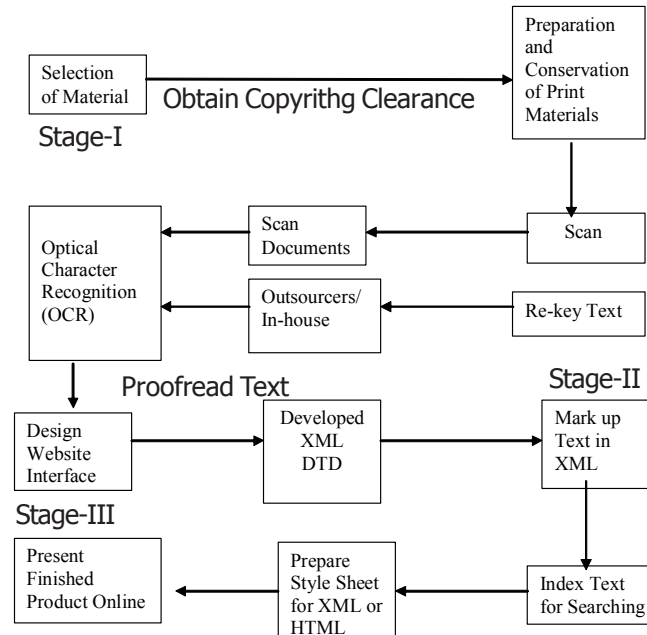


Figure 1 : Digitising Process

Source: Rakesh Prasad. *Preservation of Library Resources in Digital Environment*. Univ. News, 2006. 44(11). p.9.

9.1 Post-processing

The scanned file needs some processing in order to make them suitable for access through a network which involves following activities-

- a. conversion to a suitable file format with/without compression
- b. the creation of an index using metadata and/or the full content
- c. the creation of an interface for searching and/or browsing the metadata.

10 Access to Digitised Information

Access to the digitised information can be provided in a number of ways. The most common approaches adopted in the Digital Libraries are searching and browsing.

Searching may be based on the words/contents of the digitised resources which is enabled by indexing the full texts of the digitised materials or indexing according to some metadata structure, such as, author, title or descriptor.

In Browsing mode, users can glance through a list or hierarchy for the required digitised information resource. The categories of the may be based on certain keys, such as, authors or titles, or it may be based on subject classification approaches (Chowdhury and Chowdhury, 2003).

11. The ETD Scenario

11.1 International Scenario

Some of the International initiatives are discussed here.

Networked Digital Library of Theses and Dissertations (NDLTD)

NDLTD leads a movement of ETDs worldwide supported by the Virginia Tech University, USA. It is an open federation consisting 174 member universities and research institutions from all over the world. It takes an initiative for the universities, libraries, research students in order to support authoring, indexing, archiving, disseminating and retrieval of ETDs worldwide.

NDLTD's vision is to –

- a. Increase the availability of research documents
- b. Preserve it electronically
- c. Encourage and support the universities to unlock their information resources

with an advanced Digital Library technology by sharing of experiences, tools, technologies and knowledge.

UNESCO Initiative

UNESCO's one of the main goal is "free exchange of ideas and knowledge" is to redefine universal access to information and the minimum level of service to be provided to information users by the public sectors (Singh, 2005). The UNESCO has launched a project for development of an international framework for creation of ETDs. This project aims to contribute to enhancing the production, access and archiving of scientific information by using the possibilities of new technologies. The specific objectives of UNESCO project are-

- establishing and disseminating guidelines work flow model and best practice
- establishing a model training programme for project managers responsible for ETD programmes
- carrying out training courses and pilot projects (Vijaykumar, 2004).

Worldwide, other notable ETD initiatives strengthening and augmenting higher education and academic research are:

Table :2, ETD Initiaves Projects

Name of the Country	Name of the Project	Available Site
Australia	Australian Digital Theses Project	http://adt.caul.edu.au
Canada	Canadian Theses Project	http://www.nlc.ca
France	University Lumiere Digital theses Project	http://www.uni-lyon2.fr
Germany	Die Deutsche Bibliothek(The National Library of Germany)	http://www.dissonline.de
USA:	1. Virginia Tech 2. California Institute of Technology	http://scholar.lib.vt.edu/theses http://library.caltech.edu/etd

11.2 Indian Scenario

Endorsement by Government of India, National Task Force on Information Technology and Software Development. IT Action plan III Content Development Policy- (item 81) states –" It will be made mandatory for all the Universities in the country to host every dissertation/thesis submitted for research degrees on a designated website"

In India, there are many initiatives have taken by the institutions/organisations for the ETD research and development. These initiatives help to establish even consortium approach among the libraries. Some are summarised below:

11.2.1 Vidyaniidhi (Digital Library and E-scholarship Portal)

Vidyaniidhi (meaning "Treasure of Knowledge" in Sanskrit), a nation-wide effort on theses and dissertations, in collaboration with NISSAT, DSIR, GOI and Ford Foundations and Microsoft India has initiated a project to facilitate the creation archiving and accessing of the full-text doctoral theses from various universities. It is envisioned to evolve as a national repository and a consortium for e-

theses through participation and partnership with universities academic institutions and other stake holders (Natarajan, 2005). It host 500 theses in full-text and 85,000 bibliographic records of theses submitted to the universities in India (Vijaykumar, 2004). Vidyanidhi is now a member of the Networked Digital Library of Theses and Dissertations (NDLTD). The official website of Vidyanidhi is: <http://www.vidyanidhi.org.in/asp/search/guide/html/main.html>.

Up to 2004 Vidyanidhi has evolved into an online Digital Library with a database of 60,000+ Theses (coverage from 1992-2003) and a full-text collection of 500+ Theses. The main objectives of the Vidyanidhi are

- To involve as an information infrastructure and an enabling frame work for strengthening research capabilities of Indian Universities.
- To develop organisational model and technical mechanism for creation, submission, archiving and accessing of Indian Theses
- Developing accessible digital libraries of Theses and Dissertations (ETDs)
- Training the doctoral research students in scholarly technical writing, i.e., E-publishing and ETDs
- Developing appropriate strategies for Metadata Standard and requisite software platform and other resources (Dasgupta and Banerjee, 2006).

11.2.2 Centre for Development of Advanced Computing (C-DAC):

C-DAC has developed a high performance-computing cluster PARAM Padma for Digital Library applications and initiative to store and retrieve digitised contents at various universities and information centres.

11.2.3 Information and Library Network (INF LIB NET):

INFLIBNET, an IUC of the UGC host databases of bibliographic records of Ph.D. theses submitted in various universities in India. The INFLIBNET mission is "to ensure free flow of relevant information conducive to advancement of academic community in particular and the society at large and resource sharing of knowledge and promoting creativity that new technologies provide". The new roles that are need of the hour in the digital environment is the access to ETDs. It consist of nearly 1,40,000 bibliographic records (INFLIBNET, 2001).

There are also a number of other agencies that are involved in collection, compilation and presentation of metadata of theses in India.

11.2.4 National Social Science Documentation Centre (NASSDOC)

NASSDOC, a wing of the Indian Council of Social Science Research (ICSSR) host a library for Social Scientists with Ph.D. theses in Social Sciences. Currently NASSDOC host nearly 4,000 theses in Social Science in its collection.

11.2.5 Developing Library Network (DEL NET)

DELNET also host a database of bibliographic records of Ph. D. theses submitted in various universities in India. It consists of nearly 5,000 bibliographic records.

11.2.6 Association of Indian Universities

AIU publishes list of theses awarded in various universities and also a number of bibliographies on theses submitted to the Indian universities in various disciplines in its weekly publication called "University News". It will be far better if these are digitised.

12. Conclusion and Suggestions

ETDs eliminate the limitations of print form of cost more of reproduce, time-consuming to publish, and the inter-library loan programmes. Thus, many theses and dissertations are unread in libraries. To success the digital access to the global research output ETD, Researchers should submit ETD in the universities along with the print form and it should be mandatory to recommend a draft about ETDs Documentation which should include-

- a resource targeted to the research students who are writing theses and dissertations,
- for faculty who want to mentor ETD authors,
- for research administrators to initiate ETD programmes,
- converting a theses or dissertation to PDF file,
- submitting the approved e-theses,
- how the files would be transmitted to the university libraries,
- and the users can access, read print or download them (Weisser, 1997).

I think that if the digital information storage, delivery and access should be a motto in keeping with the cherished dream of Rabindra Nath Tagore for storing universal knowledge for the benefit of the students, researchers in particular and the community at large.

13. Acknowledgement

I am grateful to the Librarian, Central Library, Visva-Bharati, West Bengal to allow me to publish and present this paper to the PLANNER-2006 and also grateful to the Incharge, Gitanjali Net, INFLIBNET Center, Visva-Bharati for necessary help.

Reference

1. Chowdhury, G.G. and Chowdhury, Sudatta (2003). Introduction to digital libraries. London: Facet Publishing. pp.103-119
2. Dasgupta, Ahbhijit and Banerjee, Manika (2006). Digital Library of electronic theses amongst the government funded scientific research organizations in Kolkata: state-of-the art: an analytical study. In: S.S.Bandyopadhyay, B.Sutradhar and S.S.Pathak (eds.). Information Management in Digital Libraries. NCIMDiL Conference held at Kharagpur, August 2-4, 2006.pp.205-216

3. Dougherty, Richard M. and Hughrs, Carol (1991). Report of the task force on the electronic journal. Available (Online) <www.nal.usda.gov/services_and_products/other_nal_products/eii/bib.html> website visited on August 5, 2006
4. Finter, Yale (2003). Electronic theses and dissertations. Available (Online) <http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v003/3.2fineman.html> , website visited on August 9, 2006
5. INFLIBNET Newsletter (2001). 7(2). Available (Online) <<http://www.inflibnet.ac.in/newsletter.html>>
6. Nair, R.Raman (2002). Internet for information services. New Delhi: Ess Ess Pub. pp.1-33
7. Natarajan, M. (2005). Cyber infrastructure: an opportunity for education and implications for research libraries. *Annals of Library and Information Studies*, 52(4). pp.147-153
8. Prasad, Rakesh (2006). Preservation of library resources in digital environment. *Univ. News*, 44(11). pp.6-10
RGUL (2006). The Robert Gordon University Library Service: Electronic theses. Available (Online) www2.rgu.ac.uk/library/etds/why.html. website visited on August 13, 2006.
10. Siddiqui, M.A.(1997). The use of information technology in academic libraries in Saudi Arabia. *Journal of Librarianship and Information Science*, 29(4). P.195
11. Singh, Anil (2005). Electronic theses and dissertations: an indispensable resource for global dissemination of research output. *International Library Movement*, 27(4) pp.205-212.
12. Sinha, Atanu Kumar (2006). Role of digital libraries in electronic literacy in the age of e-publishing. In: S.S.Bandyopadhyay, B.Sutradhar and S.K Pathak (eds.). *Information Management in Digital Libraries*. Proceedings of NCIMDiL held at Kharagpur, August 2-4, 2006.
13. Vijaykumar, J.K., Murthy, T.A.V. and Khan, M.T.M. (2004). Electronic theses and dissertations for Indian universities: a framework. Proceedings of the 2nd PLANER- 2004. Imphal, December 4-5, 2004. pp.65-70.
14. Weisser, Christian; Barker, John and Walker, Janicr R. (1997). Electronic theses and dissertations: problems and possibilities. Available (Online) <<http://www.cas.usf.edu/english/walker/papers/etds/etds.html#concl>> , Last modified on March 30, 1997. website visited on August 13, 2006
15. Zhang, A. (1999). Multimedia file formats on the internet: a beginner's guide for PC users. Available (Online) <<http://www.lib.rochester.edu/multimed/contents.htm>> website visited on August 16, 2006.