
ELECTRONIC THESES AND DISSERTATIONS [ETDS] IN TEA SCIENCE AND TECHNOLOGY - PROPOSAL FOR A NETWORKED REPOSITORY OF TEA ETDS

Govind Joshi

Abstract

Theses and dissertations are important knowledge resource created by graduate students in universities or institutes of higher learning. Due to lack of bibliographic control, and restricted access, these grey literature collect dust on shelves of libraries. Hence these are lost to prospective researchers and students to further their studies etc. Although many national and international efforts are being made to collect, collate and distribute theses and dissertations in many forms, none is comprehensive. The present trend is to prepare the theses in electronic form to facilitate easy online access to metadata, abstract and table of content or full text, the theses become OAI-PMH compliant and can be searched on internet from anywhere. Depending upon access restrictions the theses can be harvested and stored in any institutional repositories. Indian scenario of theses and dissertations as knowledge resources, and maiden efforts by Tocklai Library in creation of Institutional Repository [IR] of all theses and dissertations in tea science and technology including other aspects are discussed. Proposes for a networked repository of TeaETDS among possible stakeholders in the plan.

Keywords : ETD, Tea Industry, Institutional Repository, LOCKSS, OAI-PMH

1. Introduction

Masters' theses and Doctoral dissertations are fruits of study and research undertaken by graduates in universities and research institutes. These are intellectual capital of any organization. These are important resources used by others in creating new resources. Although it is not easy to ascertain the total number theses and dissertations] currently available in numerous libraries in India or around the world, according to one estimate the number of doctoral theses alone will be more than 2,00,000 (reference required). The number of masters' theses will be much higher. There are more than 350 universities and 17000 colleges in India where nearly 60,000 (reference) research scholars are enrolled in a year pursuing their masters/doctoral degrees. In addition to universities and colleges, a number of apex bodies of higher education and research organizations such as the CSIR, ICAR, ICMR, ICHR, ICSSR, DBT, DST under the Government of India and also some private bodies offer research fellowships to scholars to pursue research activity which result in creation of theses. At any given point of time the CSIR will be supporting some 7600 research fellows under different schemes while recruiting some 1200 research fellows annually. That means each year thousands of doctoral and masters theses are produced in India alone. But unfortunately there is no single

agency which collects the information comprehensively and exhaustively and makes them available to info seekers. What is available is the list of theses submitted to various universities published in the weekly editions of the *University News* – a publication of the Association of Indian Universities of the UGC. INFLIBNET, VIDYANIDHI, DELNET and INDEST are some Indian initiatives who are creating databases of doctoral theses for the benefit of the scholarly communities.

INFLIBNET	153,206+ (175000) bibliographic records	- Openly Accessible
VIDYANIDHI	50,000+ bibliographic records	- Openly Accessible
	4151 Full Text theses in DSpace	- Openly Accessible
DELNET	44,304 bibliographic records	- For Members only

However there is a lot of overlap in the holdings of the above databases. Still thousands of Ph. D. theses are lying on the shelves of university libraries as grey literature and therefore lost to active researchers outside the university library. Access restrictions imposed by the university is another factor.

2. Electronic Theses and Dissertations [ETD]

Electronic Theses and Dissertations, simply put, are theses and dissertations in electronic/digital form. ETDs are the product of the digital era. It is an extension of digital library. An ET/ED is prepared using any word processing software like Microsoft Word, Word2003, Word2003 Upgrade, Office2003, WordPerfect Office12, Microsoft Works2005, AbiSource -AbiWord etc. The advantages of preparing a thesis/dissertation in electronic form are many. Multimedia files can be added to the ET/ED to make them more dynamic. An ETD can be accessed on the internet simultaneously by many. It is available 24 x 7 and is permanently preserved. Fig. 1 shows the life cycle of an ET/ED.

3. Beginning of ETD Era and International Initiatives

The ETD program was started at the Virginia Polytechnic Institute and State University [Virginia Tech] USA in 1996. From then onwards Graduate students were required to submit their theses / dissertations in electronic form only. However the concept of Electronic Theses and Dissertations was first discussed openly at a meeting at Ann Arbor, USA, in 1986 by UMI-the world leader in theses and dissertations database. Virginia Tech, which attended the meeting, funded a project on ETD at its campus. Today it has one of the largest collection -10,145 electronic theses and dissertations in its database. A number of national and international efforts have been made since then. Networked Digital Library of Theses and Dissertations [NDLTD] is an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper based theses and dissertations [1]. The NDLTD has more than 190 members including 160 universities, 6 consortia, 24 other institutions around the world.

4. Other International Initiatives

4.1 Theses Canada

Theses Canada is a portal the objective of which is to acquire and preserve a comprehensive collection of Canadian theses at Library and Archives Canada (LAC), to provide access to this valuable research within Canada and throughout the world. It is a voluntary program. There are about 2,75,000 masters and doctoral theses in the collection.

4.2 EThoS [Electronic Theses Online Service]

A program funded by the Joint Information Systems Committee (JISC) and Consortium of Research Libraries [CURL]. It is based at the British Library. The mission statement of EThoS says that the aim of the EThoS project was to deliver a fully operational, easily scaleable and financially viable prototype UK online electronic theses service, and supporting infrastructure. This service will enable students, researchers, and in fact anyone else with an interest in post graduate research theses material, to search the EThoS and to access, from the desktop, the full text, in secure format, of electronically stored theses, following selection. It will also enable UK Higher Education institutions, in partnership with the British Library, to promote their post graduate research and increase usage of their theses output. Metadata of over 5,00,000 records are available.

4.3 Australian Digital Theses Program [ADT Program]

The ADT program aims to establish a distributed database of ETD produced by the postgraduate research students at Australian universities. Most of the ETDs are available worldwide via the web. The sole idea behind the program is to provide access to, and promote Australian research to the international community. The initial project was funded by an Australian Research Council (ARC) - Research Infrastructure Equipment and Facilities (RIEF) Scheme grant (1997-1998). ADT is the initiative of 7 Australian Universities and the Council of Australian University Librarians [CAUL]. The ADT became officially the Australasian Digital Theses Program in September, 2005, when CAUL adopted the revised business plan which includes the New Zealand Universities as full members.

4.4 Lithuanian Academic Electronic Library [eLABa] : (format)

eLABa is a UNESCO funded program in which 14 Lithuanian Universities are taking part. The activities are coordinated by Kaunas University of Technology. All the universities have submitted 4420 ETDs. Although the program is based on Fedora OSS - full text ETDs are accessible with authorization only. The ETDs can be searched in NDLTD Union catalogue[2].

4.5. University of Sao Paulo Digital Library of Theses and Dissertations [USP DLTD]

The program was inaugurated in 2001 at the University of Sao Paulo [USP], Brazil. In 2006 the USP produced some 2300 Doctoral theses and 3200 Masters Theses but most of it was not accessible to public. Now the USP has made it obligatory that students and graduates submit their theses in the USP DLTD. Currently 7000 documents are openly accessible to more than 130 countries [3].

There are a number of other international initiatives in ETDs. University libraries and research and academic institutes have established their own IRs. ETD of Hong Kong University, West Virginia University Electronic Theses and Dissertation Program, Chinese and Japanese ETD databases – Academic Research Database Repository [www.dbr.nii.ac.jp] etc.

5. ETDs –Indian Scenario

Although VIDYANIDHI, INFLIBNET and DELNET have created some databases of theses and dissertations and also electronic theses and databases, with some exceptions, the ETD culture is still in embryonic form in India. Many research institutes and IITs have created IRs of their research results using Open Source Software [OSS] like DSpace, Green Stone Digital Library [GSDL] while Eprints and Fedora are also being experimented with. These IRs have in addition to other literature, theses and dissertations also. The notable ones are VIDYANIDHI at University of Mysore [4151 FullText ETDs], Dspace@IIET [259], ETD@IISc [229], Dspace@NCL [209], OAR@IIAP [83], IITD [30], IITK [9300], IITM [3500], University of Hyderabad [4000] etc.

INDEST [Indian National Digital Library of Engineering Science and Technology] Digital Theses Archive [INDEST-DTA] – is an ambitious project in which the core members of the INDEST Consortium proposed to establish an archive of digital theses and dissertations in which all the Indian Institute of Technology and the Indian Institute of Science will be the major stakeholders. In addition to creation of a central repository of ETDs of INDEST stakeholders, the proposal also includes creation of distributed national and regional archives, retro-conversion of existing theses to electronic form, facilitate submission of theses in electronic form etc. [4].

5.1 CSIR e-Thesism

As already seen the Council of Scientific and Industrial Research [CSIR] –the second largest network of laboratories recruits 1200 research fellows annually who carry out research in more than 250 institutes, universities and other academic organizations. Theses and dissertations produced by these research fellows are being held in the CSIR e-Theses Database [5].

5.2 NASSDOC Doctoral Theses Repository

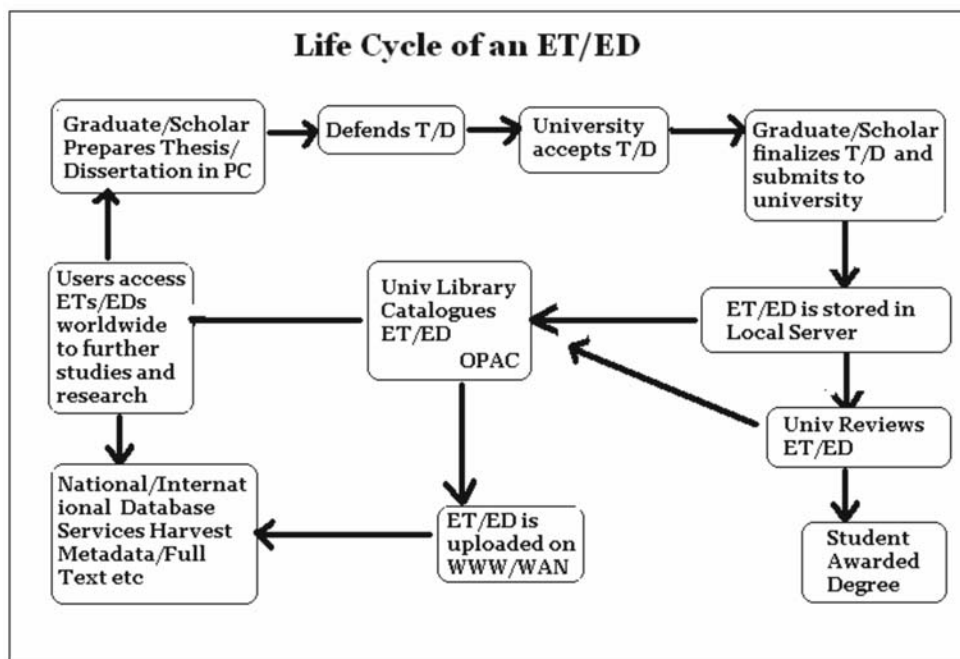
The National Social Science Documentation Centre – a constituent of the Indian Council of Social Science Research has been collecting theses in the area of Social Sciences and is in the process of digitizing. It is one of the stakeholders in the Vidyanidhi project. There are nearly 5000 (4000) doctoral theses in its database[5].

Although there had been no definite policy regarding ETD creation/submission students and researchers at the IITs, IISc, IIAP, NCL, NIO etc have been creating their theses document in computer. In 2005 the University Grants Commission [UGC] drafted a national policy framework entitled "UGC [Submission of Metadata and Full Text of Doctoral Theses in Electronic Format] Regulations, 2005". This Regulation proposed two sets of planned actions:

1. Creation of Indian National Theses Database [INTEED] : INTEED is an online bibliographic database to which every PhD student submits metadata sets of his/her PhD theses after submitting the thesis to the university. The INTEED will be freely accessible online.
2. Submission of PhD theses in Electronic Form : Research scholar submits thesis in both electronic form alongside hard copy to the university the electronic form is stored in an ETD repository which is OAI-PMH compliant. The university will setup committees to take care of copyright, plagiarism, access policy etc. [6]

Electronic thesis /dissertation is either born digital or converted from hard copy by scanning and converting to PDF [Portable Document Form] using Adobe Acrobat software, images to .pdf [for line drawings, maps etc], .jpg [for photographs, .gif [for drawings], .tiff . video files are converted to MPEG [.mpg] and audio to WAV [.wav]. [Fig. 1]. When an institute decides to create a digital repository of its theses and dissertations including other in-house created literature. Policy decisions should be made at appropriate levels. It will be necessary to constitute committees to guide and coordinate different activities involved. The entire program should be in phased manner. While first phase may concentrate on born digital thesis/dissertations, the second phase may consider retro-conversion of traditional theses/dissertations. Retro-conversion of theses is a challenging job and before launching such a program it is essential that proper policy decisions are made in relation to IT infrastructure, scanning old documents, quality check, creation of appropriate file type, metadata standards i.e., DCMI or AACR2, authors permission in relation to copyright, accessibility etc, hosting the database on local server or WAN, budget, human resource and whether or not to outsource the program.

Fig. 1. Shows the life cycle of an ETD



Born digital ETDs are those which are created using Personal Computers, and any recommended software such as Microsoft Word for creating text, graphs, line drawings etc, Multimedia files such as audio, video and photographs, flash etc can be added to ETDs prepared in computers. *LaTeX* and Adobe In Design are also being used. Although graduates and researchers in universities outside India are required to create and submit their theses /dissertations in electronic form only, the same cannot be expected in Indian universities unless the UGC Regulations 2005 [6] with regard to ETD policy frameworks are implemented. The UGC Regulations with regard to submission of PhD thesis in electronic form should be equally applicable to all other research and academic institutes which support research scholars.

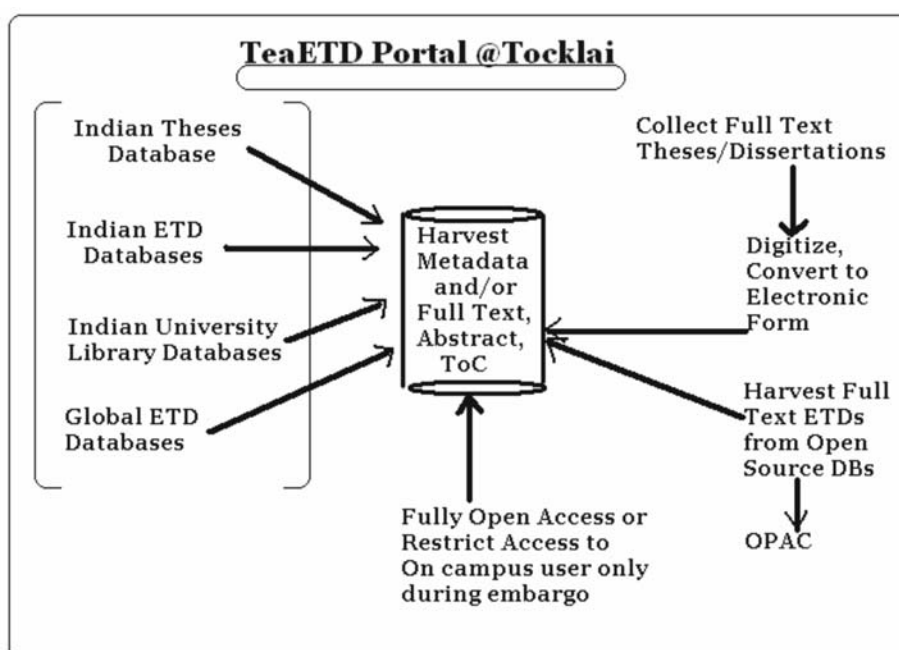
Most important aspect of ETD is the need to educate the students and research scholars in creating their documents in electronic form. A number of useful guides and tutorials, websites etc are available to the graduate students. UNESCO ETD Guide website at <http://etdguide.org/> is one good example. There is also a need to create awareness among the students, researchers and the academia regarding IPR and copyright issues and software for combating plagiarism.

5.3 TeaETD Portal@Tocklai

Most of the databases including the above three are not subject specific databases. These are like

"mixed fruit basket" one has to pick and choose from a large collection. Research and Development organizations concentrating on single crop or item have built up their subject based IRs including theses and dissertations. At Tocklai Experimental Station [TES] –the world's oldest and largest tea research institute we are building a collection of theses and dissertations in tea science and technology. It would serve as a portal for all tea related theses and dissertations. The Fig. 2 below is indicative of flow of work at the TeaETD Portal at Tocklai.

Fig. 2. Workflow at the TeaETD Portal at Tocklai



During the past five decades Tocklai alone has produced over 60 Doctoral theses in many aspects of tea science. Although many of the theses are not available physically at the central library efforts are being made to procure them from concerned universities. The Gauhati University which has an equal number of PhD theses has agreed to supply copies of the same. Similar arrangement is being made with other universities such as: Dibrugarh University, North Eastern Hill University, Jawaharlal Nehru University, University of North Bengal etc for Indian Ph.D. theses in tea science and technology.

Foreign doctoral theses are available from individual university libraries as well as national libraries, ETD network services. A bibliographic database of doctoral theses in tea science created at Tocklai has information on more than 600 doctoral theses in tea science. [Fig. 3]. Masters theses are not included at the present in view of the large number of the same. We are also arranging to procure

the theses from foreign universities and other agencies. Full Text and/or Abstracts and Table of Content of EIDs which are freely available on the internet are harvested and organized in a database which can be accessed within Tocklai campus in Local Area Network [LAN]. About 40 such EIDs are currently available in the database. This is a maiden effort in the North East India. [Fig. 4].

Title Effect of Chinese green tea on diet induced hypercholesterolemia and arteriosclerosis in rats.

Creator [Researcher]	Yang, Tat-chi Teddy
Degree Grantor [University] Advisor/Guide	Hong Kong University
Date [Year of Award]	1998
Subject [Keyword]	Pharmacology
Language	English
Bibliographical Level (Database) Persistent link	Ph.D. vii 259pages. Complete Text can be downloaded with permission only. Full Text in pdf form available. http://sunzi.lib.hku.hk/hkuto/record/B31237939
Last updated Date	2-7-2007
University/Institute	Email Request
Address	University of Hong Kong Pokfulam Road, Hong Kong, R.O.C.
Librarian	Dr.Tony Ferguson (University Librarian)
Phone/Fax	(852) 2859-2203 Fax : (852) 2559-5045 / (852) 2517-4615
Email	hkulref@hkulcc.hku.hk
Website	http://lib.hku.hk

Fig. 3. Bibliographic Database of Doctoral Theses in Tea Science at Tocklai 74/2007

Name :	Chen, Jie, Jack				
Title :	Growth inhibition effects of green tea and epigallocatechin gallate in bladder tumors				
Persistent Link:	http://sunzi.lib.hku.hk/hkuto/record/B29798991				
	<table border="1"> <tr> <td>LOCATION</td> <td>CALL NO.</td> </tr> <tr> <td>Special Collections</td> <td>Ph.D.04 C17</td> </tr> </table>	LOCATION	CALL NO.	Special Collections	Ph.D.04 C17
LOCATION	CALL NO.				
Special Collections	Ph.D.04 C17				
Year :	2004				
Pagination :	xx, 175 leaves : ill. ; 30 cm.				
Language :	English				
Dept/Program :	Pharmacology				
Degree :	Doctor of Philosophy				
LC Subject :	<u>Catechin.</u> <u>Green tea - Therapeutic use.</u> <u>Bladder - Cancer - Genetic aspects.</u>				
Abstract :	PDF QCRd text				

Fig. 4. Screenshot of Metadata harvested from OPAC - internet resource

The Open Access Initiative [OAI] has come as a boon to the LIS profession especially when cost of subscribing to high impact journals, databases and other documents essential to scholarly communities is skyrocketing. Thousands of journals and databases are now openly and freely accessible on the internet. The university libraries around the world are also throwing open their resources to other communities. The result is institutional repositories [IR] of their resources. EIDs form a major chunk of the IRs. The EIDs are available in many IRs either immediately after acceptance by the universities if there is no author's embargo or accessibility is limited to Abstract/ToC till the embargo is lifted after 2 or 3 years. Fig. 5 shows the result of a EID search of University of Hong Kong Library System.

Thousands of EIDs are available in the IRs of university libraries in South East Asia. But a large number of the documents are in languages other than English and therefore lost to the English speaking community.

This clearly shows although the EIDs allow the flexibility of creating documents in any preferred language, the same are rendered almost useless to others. EID- IRs should be able to provide automatic translation of documents is necessary.

Title Effect of green tea on diet-induced obesity in rats

Creator	Mai, Dan, Karen
Creator	¥•9N
Subject	Green tea - Therapeutic use.
Subject	Obesity.
Subject	Rats - Physiology.
Publisher	University of Hong Kong (Pokfulam Road, Hong Kong)
Date	2001
Type	Thesis or Dissertation
Format	print
Identifier	http://sunzi.lib.hku.hk/hkuto/view/E23295090/tc.pdf
Source	http://sunzi.lib.hku.hk/hkuto/record/E23295090
Language	eng
Rights	unrestricted
Rights	The author retains all proprietary rights, (such as patent rights) and the right to use in future works.

Fig. 5. Screenshot of a Search Result Page

6. Tea And ETDs

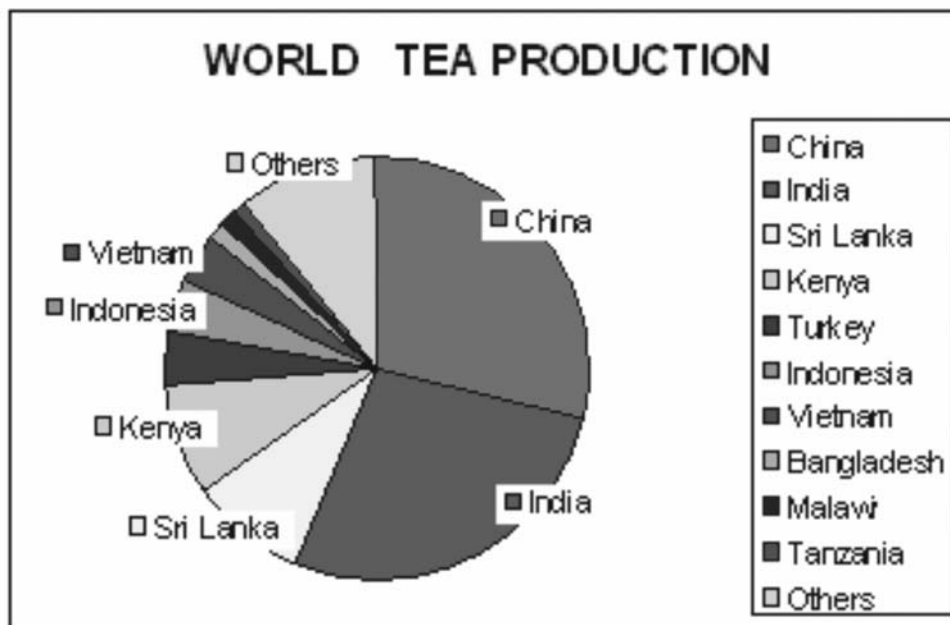
Tea - *Chay* - *Chaya* - *Ça* - the world's most popular drink is also the oldest beverage and health drink. China, India, Sri Lanka and Kenya are the world's four highest tea producing countries in that order. Table 1 and Fig. 6 shows the top ten tea producing countries and their production. Tea is grown in more than 50 countries around the world in a tea belt which spreads 45°N and 35° S of the equator. Tea industry around the world is one of the most labor intensive industries. India contributes up to 30% of the world tea production and Assam contributes a whopping 60% of India's total production.

Table 1. Top Ten Tea Producing Countries of the world [figures in M.Kg]

Country/Year	2004	2005	2006
China	835.23	934.86	1020.00
India	892.96	927.98	955.91
Sri Lanka	308.09	317.20	310.82
Kenya	324.61	328.58	310.61
Turkey	165.82	135.00	142.00
Indonesia	164.82	165.85	139.76
Vietnam	97.00	112.00	132.00
Bangladesh	55.63	58.62	53.26
Malawi	50.09	37.98	45.01
Tanzania	30.69	30.36	31.35
Others	390.83	387.22	382.76
Total	3,314.95	3,435.65	3,523.48

Source : Dr SS Baghel, VC AAU "R & D Focus on Tea in NE India" PowerPoint Presentation, 22-9-2007

Fig 6. Shows Percentile Production of Tea



7. A Proposal for Networked ETDs in Tea Science and Technology

There are four research institutes actively engaged in research and development activity in tea science and technology in India. These are : Tocklai Experimental Station [TRA] , Tea Research Institute [UPASI] , Darjeeling Tea Research and Development Centre [Tea Board] , Institute of Himalayan Bioresource Technology [CSIR] . A good number of universities are also doing fundamental and applied research in many aspects of tea science and technology. The result is production of a large volume of high quality literature including theses and dissertations in this subject. About 250 PhD theses are listed in our bibliographic database and the figure is not exhaustive. Since the theses and dissertations are scattered in a number of institutes and / or university libraries and there being no centralized agency or coordinated effort to serve all tea related theses/dissertations we at Tocklai propose to create a central warehouse of all tea theses/dissertations . All the above named Tea Research Institutes can be possible stakeholders in the program. [Fig 7] Using Open Source Software like LOCKSS [Lots Of Copies Keep Stuff Safe] copies of the ETDs can be made available in all the stakeholder institutes. LOCKSS, founded in 1998 is a Stanford University program provides for decentralized digital preservation of documents and also continued access to scholarly publications long into the future.

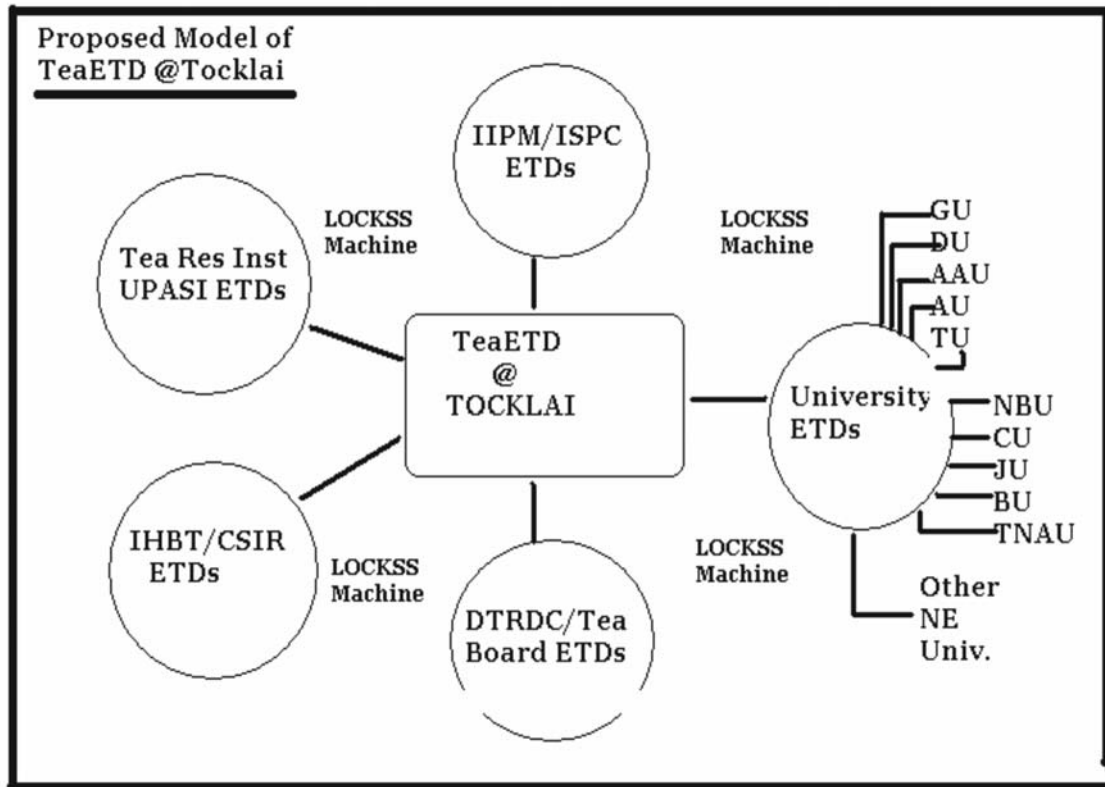
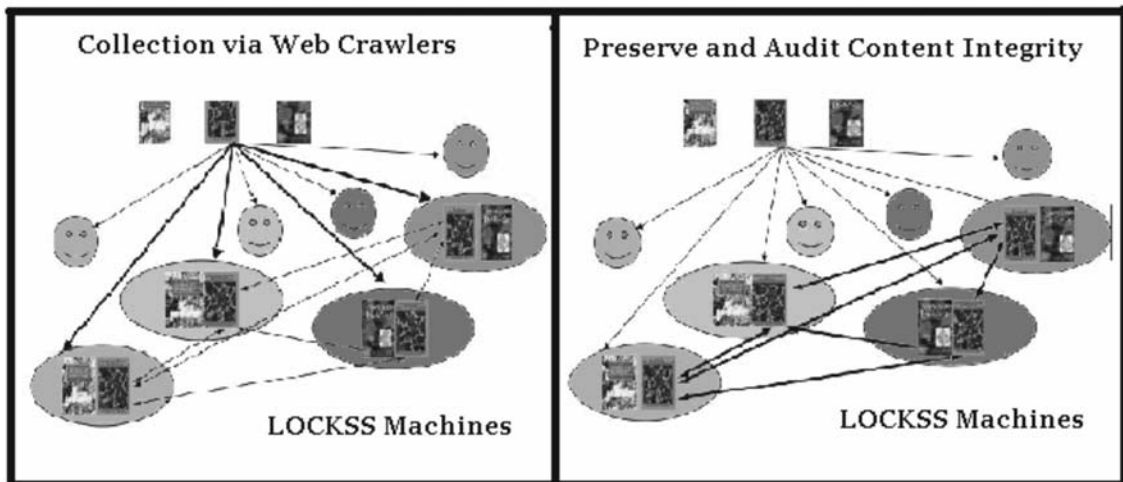


Fig. 7. Proposed Model of TeaETD@Tocklai

Scholarly communities, administrators and the librarians around the world desire and require that the research output in the form of theses and dissertations be preserved and made accessible for long period of time. Physical preservation of theses and dissertations are no more advocated for obvious reasons. However, inspite of many challenges, the information contained in these documents can be preserved in "bits and bytes". LOCKSS software [<http://locks.stanford.edu>] supports replication, distribution and reliable preservation of content. It has been reported that NDLID and six universities have collaborated in a program using LOCKSS and OAI-PMH for international preservation of ETDs [7].



How LOCKSS Works

Source : <http://www.lockss.org>

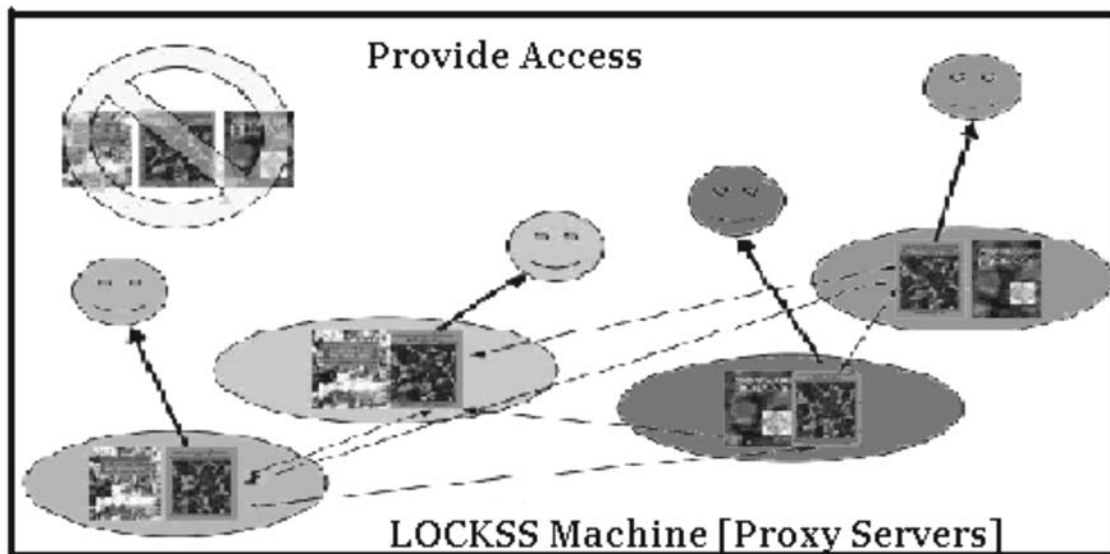


Fig. 8. The technology of LOCKSS

8. Conclusion

Theses and dissertations are important resources created by laborious and time consuming researches by graduates and scholars. These often serve as starting material to further development

in science. Unfortunately a large volume of such resource is hidden away from potential users. Although there are a number of international efforts to establish networked theses and dissertations problems like IPR and Copy Right issues, plagiarism, open access, access to full text or abstracts and table of contents only are some of the issues that need to be addressed. Theses and dissertations in electronic form [ETD] have many added advantages like online availability, accessibility to unlimited number of people on 24x7 basis, they can secure authors greater recognition etc. EID culture is still in infancy in India although elsewhere graduate students are quite knowledgeable about ET/ED preparation. Vidyanidhi – a University of Mysore project has been doing pioneering job in creation of EID portal and educating the students. INFLIBNET, INDEST, CSIR, DELNET, NASSDOC, TES are other organizations creating general as well as specialized Databases of theses and dissertations including electronic theses and dissertations. The Indian scenario of EID service can pick up momentum only if the UGC Regulations, 2005 in respect of electronic submission of theses and dissertations, are implemented forthwith. With the availability of technology and software the myth about security, IPR and copy right issues, and plagiarism can be dispelled.

References

- 1 NDLTD. Networked Digital Library of Theses and Dissertations. [http:// www.ndltd.org/about.en.html](http://www.ndltd.org/about.en.html)
- 2 Kuciucas, V.; Streimikis, A; Franckevicius, A; Targamadze, A. [2007]. EID discovery services in Lithuania. In : EID 2007 : Added values to e-theses. Uppsala, 13th-16th June 2007. Paper. Pdf. <http://epc.ub.uu.se/etd2007/files/papers/paper-04.pdf>. [Accessed on 11-Oct 2007].
- 3 Kondo, RT; Lirani, M.L.R.; Zani, V A T; Traina, C Jr.; Masiero, P C. [2007]. Using Ajax to integrate institutional electronic theses and dissertation repository with corporate systems at University of Sao Paulo. In : EID 2007 : Added values to e-theses. Uppsala, 13th-16th June 2007. Paper. Pdf. <http://epc.ub.uu.se/etd2007/files/papers/paper.pdf>. [Accessed on 11-Oct 2007].
- 4 Venkadesan, S ; Rajashekar, T B. [2003]. Operational guidelines for INDEST digital collection of dissertations in science and engineering. http://paniit.iitd.ac.in/indest/extended/indestmeet_oct03.pdf. [9 Oct 2007]
- 5 Das, Anup Kumar; Sen, B K; Dutta, Chaitali. [2007]. EID policies, strategies and initiatives in India : A critical appraise. In : EID 2007 : Added values to e-theses. Uppsala, 13th-16th June 2007. Paper. Pdf. <http://epc.ub.uu.se/etd2007/files/papers/paper-54.pdf>. [Accessed on 9-Oct 2007].

- 6 University Grants Commission [2007]. UGC [Submission of Metadata and Full Text of Doctoral Thesis in Electronic Format] Regulations, 2005. <http://www.ugc.ac.in/newinitiatives/etd hb.pdf>.
- 7 Santhanagopalan, K; Fox, EA; McMillan, G. [2006]. A prototype for preservation and harvesting of International ETDs using LOCKSS and OAI-PMH. In 9th ETD Conference: University of Laval, Quebec City, Quebec, Canada. 7-10 June 2006. <http://www.bibl.ulaval.ca/etd2006.../.pdf> [10.10.2007].

ABOUT AUTHOR

Mr. Govind Joshi, BSc, MLibSc, Cert.French, PG-DPM, Head, Library, Planning & Information Technology Department at Tea Research Association's Tocklai Experimental Station, Jorhat, Assam, India is in the LIS profession since three decades.