
DIGITAL LIBRARY INITIATIVE IN NORTH EAST INDIA WITH SPECIAL REFERENCE TO TOCKLAI EXPERIMENTAL STATION : A CASE STUDY

Govind Joshi

Abstract

Digitization has become the buzz word in the modern Library and Information Science/ Service [LIS]. Objectives of digitization , issues related with digitization, India's initiatives are discussed. Government of India's projects, schemes presently available to libraries and individuals for preservation / conservation of rare and old books, documents and other such materials of cultural heritage are stated . Tocklai's Digitization Project is discussed. Details of born digital documents and databases which are Open Source Resources [OSR] for industrial researchers in plantation crops industry are discussed.

Keywords : Digital Library; Open Source Resource; Open Archive Initiative; Tocklai; databases; Open Access Initiative.

1. Introduction

Digital, Born Digital and Digitization are the buzz words the present Library and Information Service [LIS] professionals are familiar with. Rapid developments in computers, Information and Communications Technologies [CIT] have thrust upon the LIS professionals new and added responsibilities. Their role has undergone a metamorphosis from the traditional 'book-keeper' to 'bit-keeper' or 'bit-manager'. This is more so because of the ever growing size of the Internet and the volume [bits] of information and of course the structure of the internet which is mostly unorganised. Online information doubles every six months [1] and the total digital information that is available on the Internet is estimated at 532897 Terabytes in 2003. Researchers, student and faculty members prefer to search for information online due to the obvious reasons of speed, accuracy and accessibility from anywhere in the world.

Size of the Internet in Terabytes

- Medium 2002 terabytes
- Surface Web 167 Tb
- Deep Web 91850 Tb
- Emails [originals] 440606 Tb

Sifting the bits of information from the cyberspace is a challenge to the professional librarian. In order to satisfy the Five Laws of Library Science he has to look beyond his own library resources. He needs to cross over boundaries of his library. He needs to throw open his resources and be able to access the resources of other libraries. He needs to share the resources. This is possible only when his resources are made available online. That is where digital libraries play an important role. Digital Libraries are "libraries without walls" where one can access information 24 x 7 days and from anywhere in the world. DL breaks barriers of time, space and language and culture and facilitates resource sharing on the internet and thereby better usage of information.

The mission of conversion of traditional library resources into digital has two underlying principles firstly free access to information from anywhere from the world which is to support the global Open Access Initiative [OAI] and the Open Access Movement [OAM] [2] in the spirit of recently passed Right To Information Act [RTI Act 2005] [3] and preservation of our cultural heritage. The RTI Act defines "information" as:

Information' means any material in any form, including records, documents, memos, emails, opinions, advices, press releases, circulars. Orders, logbooks, contracts, reports, papers, samples, models, data material held in any electronic form and information relating to any private body which can be accessed by a public authority under any other law for the time being in force '[3]. Hence it is imperative that every organization , government funded , public or private make its informational resources freely available to the public or anyone seeking.

2. Digital Library of India Initiative [DLI]

The DLI initiative has its root in the 'Million Book Project' [MBP] initiated by the National Science Foundation of the USA at the Carnegie Mellon University [CMU] with India as partner. The prestigious Indian Institute of Science [IISc], Bangalore became the central point of activities. The objective of the collaborative project was to digitize one million seminal books which happens to be in the public domain or copyrighted but out of print, making them available on the web to anyone at any time anywhere in the world. The One Million Book Project was expected to be completed by December 2005. But according to the CMU website information, the current status of the Million Book Project is " As of November 2005 -

- Over 600,000 books have been scanned: 170,000 in India, 420,000 in China, and 20,000 in Egypt. Roughly 135,000 of the books are in English; the others are in Indian, Chinese, Arabic, French, or other languages. Most of the books are in the public domain, but permission has been acquired to include over 60,000 copyrighted books (about 53,000 in English and 7,000 in Indian languages).
- The books that have been scanned to date are not yet all available online, and no single site has copies of all the books that are available online.
- Twenty-two scanning centers are operating in India, including four mega-centers. Eighteen centers are running in China, including a mega-center in a free-trade zone to avoid customs delays with shipments of books from the U.S. " [www.library.cmu.edu/Libraries/MBP_FAQ.html#current]

As a part of this mega project India launched the Digital Library of India Portal [DLIP] in September 2002.[4].

Department of Culture, Government of India also launched the National Mission for Manuscripts in February 2003. The objectives of the mission are to facilitate conversion and preservation of manuscripts, and under the National Archives Directorate, old and rare books and other documents through training, financial assistance, cataloguing etc. [5]. A good number of leading libraries have initiated digitization of their special collections at the individual level. Such collections are accessible as institutional repositories only. Old and rare books, manuscripts, theses and dissertations are digitized.

- National Library, Kolkata has digitized 6600 selected books in Indian and English language [www.nlindia.org].
- Digital Library of India hosted by IISc, Bangalore [www.dli.ernet.in].
- GyanNidhi [MCIT , Govt. of India] project 1 million pages multilingual- English and 12 Indian languages [www.cdacnoida.com].
- Vidyanidhi – Theses and Dissertations [wwwvidyanidhi.org.in].

Digital Library of India [DLI] hosted by Indian Institute of Science , Bangalore in cooperation with CMU, IIIT, NSF, ERNET and MCIT for the Govt. of India and 21 participating centers. Presently the following category of documents have been digitized and are searchable with keyword/title/author.[6]

Books	News Papers	Journals	Manuscript
Rashtrapati Bhavan	Times of India	INSA	Tamil Heritage Foundation
CMU	Indian Express		
Sanskrit	The Hindu		
TTD Tirupati	Deccan Herald		
	Eenadu		
	Vaaritha		

3. Tocklai Initiatives

Our first effort to digitize old and rare publications in the Central Library was in response to the Ministry of Culture, National Archives, Govt. of India's advertisement in 2004. Unfortunately our proposal failed to get the financial support of the National Archive, New Delhi as the same was not covered under their scheme. The National Archive was more concerned with preservation of old and rare documents, manuscripts etc by way of physical preservation/conservation, microfilming and cataloguing rather than 'digitization'.

However , interestingly 'preservation by digitisation' definition also prevailed in the jargons of the Ministry of Culture, Government of India.

The silver lining in the horizon was CDAC, Kolkata with which TRA was collaborating in a joint project to establish an automated tea factory – a Model Tea Factory [MTF] offered to undertake

digitization of the Central Library Collection. CDAC – an autonomous body under Ministry of Information Technology [MoIT] has taken up digitization work in some 20 centres spread all over the country.

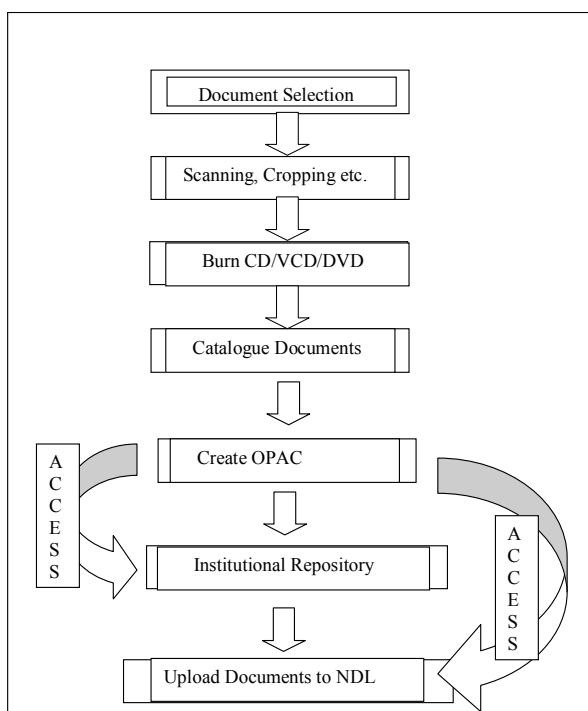
4. Tentative Status of Documents in Tocklai Central Library

The Central Library has :

■ Books	6000
■ Bound Journal volumes	30000
■ Maps	300
■ Tocklai/ITA/TRA's own publication	2000

CDAC, Kolkata has been actively engaged in digitization work since April 2005. A high speed scanner with variable platform is used for scanning the documents. Two Pentium IV computers and OCR software are being used. So far 800,000 pages have been scanned. All of Tocklai Library, Tocklai Archive, Tocklai/ITA/TRA publications have been digitized. All the digitized materials can be accessed from the institutional repository when completed.

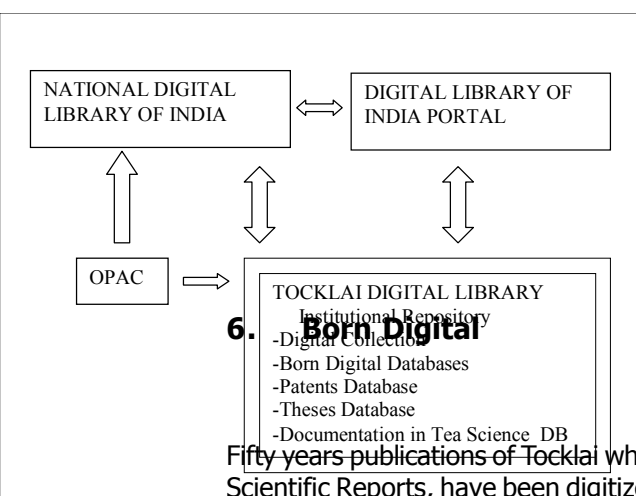
Fig. 1. Flowchart of CDAC's Digitization Work



5. Documentary Resources at Tocklai Experimental Station

Tea Research Association's Tocklai Experimental Station at Jorhat, Assam popularly known as 'Tocklai' is synonymous with tea research. Since its establishment in 1911 Tocklai has built up vast documentary resources in the area of tea science and technology. The collection includes books, journals, reports, technical literature, maps, charts, statistics and the like. Tea related publications from all over the world in all languages are procured by the Central Library. Some of the old and rare collection dates back to 1814. In addition, Tocklai has its own publications which date back to 1890s. Though of archaic value some publications are consulted by the scientists even today. All these material which are paper based represent our cultural heritage and therefore need to be digitized and preserved for future.

Fig.2. Schematic Representation of Tocklai Digital Library



Fifty years publications of Tocklai which include half-yearly research journal "Two and a Bud", Annual Scientific Reports, have been digitized and are now available in CDs. Other publication which have been digitized are Tea Encyclopedia, Planters' Handbook, Memorandum on Shade Trees and Cover Crops, Pests, Diseases and Weeds etc. All the digitized material will be available in the institutional repository.

The following in-house databases are born digital :

6.1 Documentation in Tea Science (DTS)

Documentation in Tea Science aims to collect, collate, catalogue and organize non-TRA / Tocklai publications and digitize and make it Open Source in the near future. The database has about 100,000 research papers on tea science and technology and allied subjects.

6.2 Tea Patents Database (TPDB)

Patent data is a new resource of knowledge. Patent information empowers individuals and organizations who need to innovate and stay ahead in their respective areas of specialization. Tea Patents Database created at Tocklai lists more than 2000 patents related to tea including patents in tea machinery, manufacturing methods, by-products of tea, tea based beverages, tea confectionery, tea capsules, tea cosmetics etc.

6.3 Database of Doctoral Theses Science (DDTS)

Since 1950 Tocklai alone produced some 55 doctoral theses in tea science. Hundreds of universities and research institutes all over the world are producing doctoral theses in tea. The DDTS lists more than 500 doctoral theses submitted to universities around the world. Full bibliographic details are given under each entry while some are with abstracts.

6.4 Database of Tea Standards (TStDB)

Standardization has become an inevitable aspect/process of almost everything that concerns consumers. Every stage of tea culture and production process has now been standardized. Fifteen Indian Standards and 33 ISO Standards are available today to the tea industry.

6.5 Database of Open Source Resources

Open Source Resources [OSR] are open and freely accessible. However lack of knowledge as to where to find what type of information puts even the highly knowledgeable at a disadvantage. The database lists open sources in Patents, Theses and Dissertations, Standards and Free Journals and Databases.

References

1. Size of the Internet in Terabytes. [www2.sims.berkeley.edu/.../how much infor.2003](http://www2.sims.berkeley.edu/.../how%20much%20infor.2003) [accessed on 11-9-2006]
2. Bethesda Statement on Open Access[2003, June] www.earlham.edu/~peters/fosbethesda.htm
3. Right To Information Act 2005. www.cdacnoida.in/RTI/rtiact.htm [accessed on 11-9-2006]
4. Digital Library of India Portal www.dli.iiit.ac.in/mission [accessed on 8-9-2006]
5. National Archive of India www.nationalarchive.org [accessed on 11-9-2006]
6. Digital Library of India www.new.dli.ernet.in [accessed on 11-9-2006]