

8. Digital Library of India

TDIL Programme

1. Introduction

Cultural heritage refers to implicit and explicit knowledge developed in a civilization. Archives, libraries and museums (memory institutions) deal with our major resources of cultural and scientific knowledge, both as our inheritance from the past and our legacy to the future. Knowledge based societies will move up the knowledge chain fast if knowledge of previous generations is available to them. But, knowledge in Manuscripts, palm-leaves etc is disappearing rapidly. Hence, need for digitization and universal access to cultural heritage. UNICT Task Force identifies *universal access to information as enabling solution to overcome the Sprawling Digital Divide.*

2. Benefits of Digitization

Preservation and conservation

- Digital reproductions are virtually immortal
- Reconstruction using image processing techniques
- Use of life-size print reproductions for conservation

Content management

- Access to images for researchers
- Assistance in retrospective cataloging, researching collection
- Assistance with curatorial functions
- Digital Rights Management

Dissemination and promotion

- Dissemination of heritage Knowledge via the Internet

3. What is Digital Library

Digital libraries are a form of Information Technology in which social impact matters as much as technological advancements. Future knowledge networks will rely on scalable semantics, on automatically indexing the community collections so those users can effectively search within the Interspace of a billion of repositories. Just as the transmission networks of the Internet are connected via switching machines that switch packets, the knowledge networks

of the Interspace will be connected via switching machines that switch concepts. Connectivity and training continue to be the principal barriers to integrating the global network of libraries.

4. Snapshot of International Scenario on Digitization efforts

4.1 USA:

- **Digital Libraries Initiative-phase 2 (DLI-2)** by National Science Foundation , DARPA, National Library of Medicine, Library of congress National Endowment for the Humanities.
- **DARPA's Information Management** program address (www.darpa.mil/ito/research/in) core digital library issues requiring revolutionary research technology.
- **Illinois DLI project** (<http://dli.granger.uiuc.edu>)
- **Million Book University Digital Library Programme** by Carnegie Mellon University (www.ulib.org)
- **University of California , Barkley , Digital Library project**
- **National Achieves of America** (<http://www.archives.gov/>)

4.2 Europe:

- **e-Europe initiative.** ([http:// europa.eu.int/information_society/eeurope/2005/index_en.htm](http://europa.eu.int/information_society/eeurope/2005/index_en.htm))
- **MINERVA** (Ministerial Network for Valorizing Activities in Digitization)
- This programme will create a network of European states to discuss , correlate, harmonize activities carried out in digitization of cultural and scientific content, for creating an agreed European common platform, guidelines for digitization, meta-data, long-term accessibility and preservation.

4.3 ASIA: Japan

- Kyoto University Digital Library
- Japanese Digital Museum, 2000

- NTT Digital Museum
- University of Tokyo Library
- The Asia Digital Resource Project
- Pacific Rim Digital Library

China

China Digital Library Programme (Participation of 21 Ministries of China)

It is the World's largest digital library, a database containing over 12 million documents .

Malaysia:

Digital Library network of Malaysia has a network of 21 libraries.

They are also working towards creation of cultural information database.

5. India's Initiative

5.1 Digital Library Initiative of India (DLI)

India participates in the UDL project and makes efforts to put e-content of vast Indian knowledge base in the Indian language as much as possible. The Indian activity of Digital Library (Digital Library Initiative of India) is coordinated by *Indian Institute of Science, Bangalore*.

Objectives

- To digitize and index the heritage knowledge.
- To provide a test bed that will support other research domains such as scanning techniques, optical character recognition etc
- Inter-ministerial collaboration for e-Content.
- To promote collaborative creativity and building up knowledge teams across borders.
- Involvement of Resource Centers for Indian Language Technology Solutions and COILNet Centers to digitize and web-enable contents in Indian Languages.

Digital Library of India (DLI) - Centres

Scanning centres

- **Indian Institute of Science, Bangalore** (Overall coordination)

- Anna University , Chennai, TamilNadu
- Arulmigu Kalasligam College of Engineering, Madurai, Tamil Nadu
- Goa University, Goa
- Indian Institute of Information Technology, Allahabad, Uttar Pradesh
- City and State Central Library, Hyderabad , Andra Pradesh
- Indian Institute of Information Technology, Hyderabad, Andra Pradesh
- Shanmugha Art, Science Technology & Research Academy, Thanjavore, TamilNadu
- Sri Sri Sharda Peetam, Sringeri Mutt, Sringeri, Karnataka
- Tirumala Tirupati Devasthanam, Tirupati, AndraPradesh
- Maharashtra Industrial Development Corporation, Mumbai, Maharashtra
- University of Pune , Pune, Maharashtra
- Centre for Development of Advanced Computing, Noida , Uttarpradesh
- Kanchi University, Kanchi, Tamil Nadu
- Indian Institute of AstroPhysics, Bangalore , Karnataka
- Rashtrapathi Bhavan, New Delhi
- Punjab Technical University, Punjab
- Bhartiya Jnanpith , New Delhi
- Academy of Sanskrit Research, Melkotte,
- Poonpragna Vidyapeetha, Karnataka,
- University of Hyderabad , Andhra Pradesh
- Osmania University, Hyderabad
- Salarjung Museum, Hyderabad
- Tibetan Monasteries and Literature on Jainism
- Namgyal Institute of Tibetology, Sikkim.

Progress of the UDL project in India

His Excellency President of India has launched Digital Library of India Portal (<http://www.dli.ernet.in>) during September 2003, which provides unified

information about the development of Digital Library in India.

So far, in India, **more than 100,000** books (May 2005) books have been scanned and cropped in various Indian Languages viz., English, Telugu, Tamil, Sanskrit, Kannada.

5.1 Other Digital Library Programmes in India

5.1.1 IGNCA–National Mission for Manuscripts

It is involved in the Mission mode project for digitizing knowledge base of Indian traditional knowledge embedded in the form manuscripts and creation of e-content of the same. During the year 2003-04, twenty-four manuscripts resource centres (MRC) are being established throughout the country for taking up work relating to survey, listing, cataloguing of manuscripts.

Institutions involved in the digitization initiative are:

- Archaeological Survey of India
- Indira Gandhi national Centre for the Arts
- National Library
- National Museum

- Khuda Baksh Oriental Library
- National Gallery of Modern Art
- Sangeet Natak Academy
- Centre for development of Advanced Computing (C.DAC)

5.1.2 DELNET (Developing Library Network)

It was initially sponsored by the **National Information System for Science and Technology (NISSAT)**, Department of Scientific and Industrial Research, Government of India and is currently being promoted by the, National Informatics Centre, and India International Centre, New Delhi.

5.1.3 Traditional Knowledge Digital Library (TKDL)

This project aims at developing Knowledge base on Indian Traditional medicines, dissemination of information on medicines and broadening the scope of research on Ayurveda.

5.1.4 Mobile Digital Library - Knowledge at doorsteps

To facilitate surf, access, print, and take away a book of choice anywhere and anytime

6. Technology Development under DLI : Regional Mega Centres

Performance of the Regional Mega Centres:

Name of the Center	No of Books Scanned	No of Pages Scanned (In Million)	Technologies Developed
IIT-Hyderabad	34,000	18.7	<ul style="list-style-type: none"> • OCR with workflow Management, • Summarization Tool, • Multimedia tool for Digital Library • Search engine
IIT-Allahabad	17852	5.7	<ul style="list-style-type: none"> • Hindi OCR with Gabor Filter technique • Automatic Summarization Tool
C-DACNoida	15124	6.0	Work on technology development for Digital Library has been initiated.
C-DACKolkata	~5000	1.5	Work on technology development for Digital Library has been initiated.

Infrastructure Support to Digital Library Scanning Centres and Mega-Centres

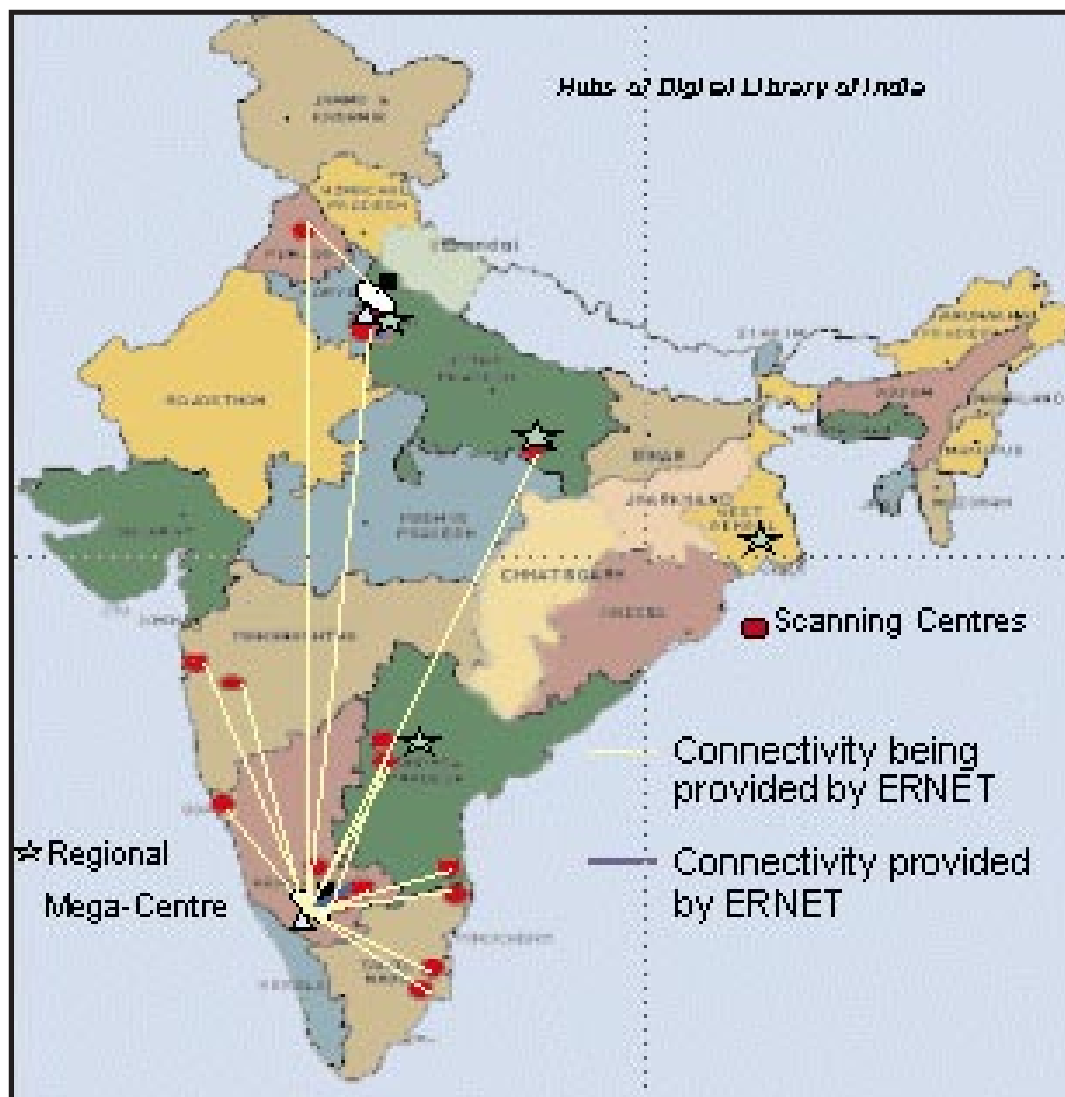
ERNET Connectivity to Digital Library Scanning Centres and Mega-Centres are being provided based on the scanning and digitization performance. A Connectivity 512 kbps is being provided to the centres who have scanned and hosted more than 5000 books, 1 Mbps link to the centres who have scanned and hosted more than 10000 books and 2 Mbps link to the centres who have scanned and hosted more than 20000 books.

7 Core Issues Involved in the Digitization initiatives

7.1 Technology Development Issues

Some of the constituent technologies that could be developed may include :

- *Integration of heterogeneous repositories* with separate semantics
- Resource *indexing and discovery*
- *Clustering* & automatic hierarchical organization of information & algorithms for automatic rating.
- *Evaluation* of information quality, genre & other properties.
- *Metadata handling* – (definition, use and its capture from objects (textual and multimedia))
- Distributed data and *metadata extraction* from multi source and its handling and
- *Interoperability* amongst resources
- *Federating* (mapping together of similar objects from different collections in a way that makes them appear as one) or mediating software



compensating for site-by-site variations search at semantic levels

- **Semantic indexing and retrieval** and Scalable Semantic
- Multilingual, multimedia multi-interface **information access** for all (physically challenged, illiterate, etc)
- Universal Seamless access via Wireless network for remote area.
- **Standards, IPR** and privacy protection
- Collaborative development and test-bed

7.2 Service Issues

- **Memory Requirement:** The culture sector would require comprehensive *technical solutions* which are sustainable, i.e. they must be robust, have a broad applicability, take into account the abilities to handle large numbers of records be consistent across categories.
- **Standardization:**
- The appropriate application of technical standards by those involved in digitisation activities is essential if the digitised collections are to be maintainable in a cost-effective way over time
- **Quality Requirement:**
- A framework is required that describes quality in terms of the underlying technical characteristics provided by individual sites or subject gateways and in terms of the quality of service provided to the users.

7.3 Legal Issues

- **Copyright:** Investigation of the legal position in relation to making digital copies of the relevant source materials as the very first stage of the digital imaging project
- **Information Security:** Taking steps to prevent **unauthorized changes** in the digital files created during the project
- **Legislation:** Ensure that the project complies with appropriate local legal deposit legislation.
- **Digital Rights Management:** Evolution of **Digital Rights Management** structure

7.4 Multilingual Issues

- Character Sets (UNICODE?)
- Representations
- Multilingual Navigation
- Translation Assistance

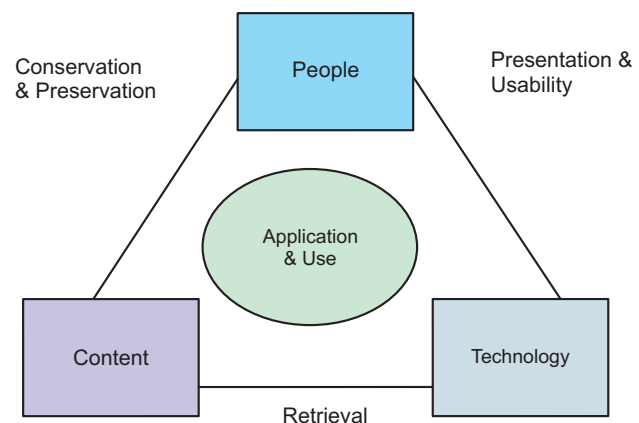
7.5 Social Issues:

- Development of user-friendly interface for better access to information by the physically challenged
- Affordable access of digitized resources.

8. Salient Operating Procedures of Digital Library:

(As evolved from International Conference on Digital Library ICDL – 2004)

Conceptual Framework for Digital Library Operations



9. Recommendations

9.1 Resources

Focus on taking the existing contents to the web and this must act as a single portal for everything that represents India. This portal could be linked as an integral part to the "global resources" for sharing with the citizens of the world.

Some of the identified resources:

- Cultural and Literary Heritage of India
- Indian History and Science
- Environment
- Tourism , Folklore and music
- Learning Materials
- Medicine

9.2 Technologies

- Optimization of the architecture between the Local storage and the Central Repository
- Development of user-friendly multiple language speech interface
- Development of Virtual Reality model for the promotion of tourism
- Evolving a language independent, human mediated technology neutral solutions

9.3 Users

- Classifications of users such as: General Population categories; Educational Categories; Professional Categories etc.
- Extension of Digital Collections online to remote and poor geographic locations and effective use of available metadata on digital content for productive retrieval and usage
- Scalable deployment to archive, search and use of very large digital collection
- Continuous user feedbacks for improvisation

9.4 Policy & Management

- *On-line Registry*
Every digital material produced in the country should be registered with the Digital Object Identifier (DOI)
- *Creation of a depository for cultural heritage*
This depository would provide authenticity and ownership as well as enable preservation of IPR
- *Copy-left provision*
This provision will enable copyright holder to give up the rights before or after a certain period.
- Copyright, lock-in period needs to be reduced (from 60 to 25 years).
- *Compulsory Licensing*

10. Future Programmes of Department of Information Technology

- Development and Deployment of on-line Multilingual Machine Translation system, OCRs in Indian Languages and Universal Dictionary Programme through TDIL programme.
- The mega-centers are focusing on the development of core technologies involved in the

development of Digital Library. Some of the core technologies that are being developed by the mega centers are:

- Upgradation of the Optical Character Recognition Systems in all Indian languages for hand-printed documents and other types of written documents
 - Development of Storage Area Networks (SAN)
 - Cross-lingual Information retrieval systems and search engines between Indian Languages and automatic summarization tools
 - Digital Preservation of Indian Knowledge base embedded in the vast Indic literature
 - Description, retrieval, object interchange and retrieval protocols
 - Metadata standardization
- Upgradation of the connectivity of the Mega-centres of scanning with 2Mb/s connectivity.
 - Establishment of **Two more regional Mega-centres** possibly in Madhya-Pradesh and University of Pune.
 - Setting up of Inter-ministerial committee to integrate Digital Library efforts in India
 - Evolution of National Mission on Digital Library.
 - Promotion of Spin-off technology from Digital Library programme.

11. AsiaNet : Asia Network for cultural resources

Content Digitization.

The foremost cultural heritage in museums, archives and libraries should be digitised according to prioritisation plans.

Universal Access.

The materials should be universally accessible to all through information networks, providing access to and promoting knowledge of Indian civilisation at the national and international levels.

Courtesy/Source:

TDIL Programme, Department of IT
Ministry of Communications & IT,
New Delhi-110003
E-Mail:tdilinfo@mit.gov.in