CHAPTER 1

Introduction

_Digital Library is where the past meets the present and creates the future..._

Dr. A.P.J. Abdul Kalam

Former President of India

1.1 Background

India is one of the emerging economies having consistent economic growth rate in last two decades due to increasing contributions from the knowledge-based service sector. Leveraging prospect of liberalization and globalization of Indian economy that was introduced during early 1990s, India has become one of the top software exporting countries. While India exports cutting edge IT products and services to the developed nations, capabilities of other knowledge-based industries are also getting strengthened in order to meet global demand of knowledge-based services as well as stay ahead in the global competition (NKC, 2006). Knowledge-based industries are human capital intensive and strengthening knowledge networks across the country, and will definitely contribute to the development of human capital (NKC, 2007). However, knowledge economy in this nascent stage could only affect a segment of Indian society, while other segment is struggling for making a decent livelihood. The use of information and communication technologies (ICT) in the cross sections of society, and more particularly in areas such as human communication, knowledge dissemination, teaching, lifelong learning and governance is also getting strengthened to match the changing economic scenario.

During the post-liberalization period in India, a significant landscape changing cum nation building exercise has been undertaken with the establishment of a huge number of new professional and technical institutions and modernization of existing educational and research institutions across the Indian states in order to produce required human capitals essential for knowledge-based industries in
India. In this process, Indian institutions have strengthened their ICT infrastructure and ICT-enabled interactive learning environment.

In contrary to India’s economic progress and prosperity, institutions providing public access to information and documentary resources to cross sections of people are on the other side of digital divide. Cultural and memory institutions in the rural, semi-urban and even urban areas across India are struggling to get operational ICT facilities due to several constraints such as lack of electricity, Internet connectivity, localized contents and trained manpower.

On the other hand, in South Asia particularly in India, large volumes of cultural heritage resources (documentary) are on the verge of extinction due to lack of preservation, non-availability, rarity and natural decay. The knowledge and wisdom lost from these cultural heritage documentary resources can cause a severe vacuum in the intellectual wealth and humanity, if we cannot preserve them at this critical juncture. The current information is also very much needed for the continuous development of the society and smooth functioning of our modern life. The information and communication technologies (ICT) act as a development enabler and intervener in uplifting quality of life through modernizing various functionalities of our life. The digital library is a product of recent innovations in the areas of ICT, where the information users can get full-text access to needed information at their workstations without delay, but with much satisfaction. Searching information from the physical documents is rather painful and time consuming if we compare it with electronic documents. The developed nations have already built up a number of digital libraries on various areas of studies. With the emergence of digital libraries, institutions across the world have strengthened their capability of serving users’ communities with adequate information resources. In India, development of digital libraries has achieved some momentum in the late 1990s and early years of the twenty-first century. The national level institutions in India took keen interest in developing institutional repositories that are now available in the cyberspace, reaching corners of the world. Later, other institutions and organizations have also started initiatives at their institutional or organizational level to disseminate new
knowledge generated within the institutions or organizations. Availability of open source software has also accelerated these initiatives in India, which has resulted in another way of disseminating scholarly literature, i.e. open access literature. In contrast to subscription-based literature, the open access literature does not have any restriction on access, and is free from any subscription fee or licensing fee (Ghosh & Das, 2007).

India is ahead of many developing countries and a few developed countries in terms of establishing a number of digital libraries or digital archives and creating digital contents for them. World communities have appraised Indian efforts, and contents of some digital libraries are regularly accessed in different parts of the world (Ghosh & Das, 2007). India also plays a crucial role in the South Asia sub-region for the development of digital libraries, which reflects the cultural diversity of the people of south Asian countries and the traditional knowledge of this region (Das, 2008).

1.2 Concepts of Digitization and Digital Library

As defined by Reitz (2008), digitization is “the process of converting data to digital format for processing by a computer. In information systems, digitization usually refers to the conversion of printed text or images (photographs, illustrations, maps, etc.) into binary signals using some kind of scanning device that enables the result to be displayed on a computer screen.” She also defines digital library as the “library in which a significant proportion of the resources are available in machine-readable format (as opposed to print or microform), accessible by means of computers”.

The digital libraries store, organize and disseminate digital contents. These contents are created either through digitization of existing printed materials and media documents, or through re-keying/re-composing of existing printed materials and media documents, or through creating new documents in digital formats. The first kind of documents is known as digitized documents, and the later kind of documents is known as born digital documents. In Indian digital
libraries both kinds of documents are available. The digitized documents are stored either in image formats or in text formats. If the original documents are available in European languages such as English, French, German and Spanish, the optical character recognition (OCR) software can automatically convert them into searchable digital text format, where qualitative OCR conversion rate is much higher. On the other hand, if the original documents are available in Indian languages such as Sanskrit, Hindi, Bengali, Oriya, Telugu and Tamil, the contents are made available either in image formats or re-keying the texts for the inclusion in the digital libraries. OCR software for Indian languages is still in the developmental or testing stage, where OCR conversion rate is much lower than acceptable rate. The full-text searching is possible in textual documents but this facility is absent in image documents.

1.2.1 Documents and Collections in Digital Library Systems

Digital library is the concept of information stored digitally and made accessible to users through digital systems and networks, but having no single location. It is, therefore, analogous to a library as a storehouse of information, but has a virtual existence in the digital spaces. Digital library is essentially a fully automated information system with all resources in digital form. Many views of digital libraries stem from what libraries currently do. Traditional libraries collect, organise, provide access to, and preserve objects in their collections. A library collection may include books, magazines, journals, theses, dissertations, manuscripts, audio-visuals, maps, etc. The flexibility of digital technology allows it to handle new kinds of object efficiently. Digital library collections can include things without direct physical analogs, such as algorithms or real time data feeds. They also may include digitized representations of what have traditionally appeared largely in museums and archives. With the rise of cost of paper publications and library storage, increasing use of computers, decreasing budgets, many libraries have to reduce their acquisition of books as well as their journal subscriptions. Documents in electronic form can become more easily available and widely used because the cost of digital storage and processing is going down.
Documents are the heart of digital libraries. Without documents there would be no digital libraries. In digital libraries, documents are not only what are stored in traditional libraries (e.g., books, journals, pictures and videos), but also include many works uncommon to those libraries, e.g., multilingual, multimedia, and structured documents (e.g., books broken into chapters, sections, subsections, figures with attached captions, colour graphics or images, attached or linked sound or video files, appendices, indexes, and 'front matter'); programs, algorithms, bulletin board archives, besides others. A document can have various representations depending on its intended use; for example, some applications require high-resolution images of documents with invisible watermarks for security purposes as well as low-resolution images for children to download from the Internet. Collections of digital library ranges from small, self-contained, and narrowly defined collections to ones spread across physical and logical spaces. One of the common requirements for a digital library is the ability to deal with distributed collections of information.

1.2.2 Evaluation of Digitization Work and Digital Library System

A digital library may be evaluated from a number of perspectives, such as collaboration pattern, system, access and usability, user interfaces, information retrieval, content and domain, services, cost and overall benefits and impact. An important issue under discussion across various communities is the set of metrics to be used for evaluating digital libraries. Selection of digital library metrics should be considered from both system-oriented and user-oriented viewpoints. From the system's perspective, we consider capacity (number of digital objects stored and number of users served simultaneously), content, transaction speed (speed of search response). From the user's perspective, we consider impacts of the system on the user (e.g., impact on patterns of association and attitudes about the digital libraries), effectiveness (relevance of the results; ability to produce a ranked list of results that are mostly relevant with best matches at the top), usability (e.g., ease of use, suitability to purpose, user's effort), interactions with the system, and user satisfaction.
In a general way, the constructs or elements for evaluation of digitization projects covered in this study are:

- Collaboration pattern for collection building;
- Collaboration pattern for resources mobilization and utilization;
- Selection of contents for digitization;
- Digitization workflow;
- Interpretation, representation and metadata;
- Access and distribution – open access versus campus-wide (closed) access;
- User interfaces – search and retrieval; and
- Integration, cooperation with other resources and libraries.

1.3 Problem Definition

As digital library system is an integrated information system, comprising metadata, multilingual texts, images, audio-visuals and multimedia data and also an efficient communication system, this system follows the system development life cycle. Beyond the post-implementation review of a digital library system, there are issues such as management of digital content, availability and accessibility of digital library services, impact on users, impact on the organization, legal constraints, retrieval efficiency, etc. need to be assessed and measured. The present study addresses the some of the necessary aspects affecting the online and digital library systems in India.

Digitization and open access projects in India serve a diverse group of users and applications, and a particular solution might not be relevant to all users and applications. Possibility of incorporating different solutions has been studied in this research work. Indian digital libraries have diverse services and products based on the mission and objectives of the host organizations. For example, proposed digital library of the National Mission for Manuscripts will provide full-text access to rare Indian manuscripts. Digital library projects of Indira Gandhi National Centre for the Arts (IGNCA) provides online access to multimedia and multilingual documents pertaining to Indian art, culture and heritage.
On the other hand, a number of institutions of higher learning in India including Indian Institute of Science (IISc), Indian Statistical Institute (ISI), Indian Institutes of Technology (IITs), Indian Institutes of Management (IIMs), National Institutes of Technology (NITs) and Indian universities are digitizing their doctoral dissertations, research reports, research papers and other technical documents for incorporating into their own institutional repositories and ETD (electronic theses and dissertations) repositories. Although few digital libraries are available in public domain as compared to the number of digitization projects undertaken across in India, Internet technologies and new communication technologies have widened its horizon to the wider communities. It is necessary to assess and measure their impact to the user communities, to review the technologies and their limitations, and to address strategic issues such as fulfillment of missions and objectives set by the digitization projects. The present study tries to incorporate some of these important aspects of the Indian digitization projects. In this research work, evaluation and assessment of existing digitization projects have been carried out on a case by case basis to critically examine technical, operational as well as managerial components of each of them.

1.4 Objectives of the Study
The main objectives of this research study are:

- To evaluate and assess the national-level digitization and digital library projects in India;
- To evaluate and assess the national-level projects in India aiming at open access to knowledge and scholarly literature;
- To evaluate and assess the digitization projects for different types of collections such as rare books, manuscripts, journal articles, documentary heritage, theses and dissertations;
- To study critically the collaboration patterns in digitization projects in India including technical and financial collaborations;
- To examine the digitization workflows in different digital library projects.
• To examine international visibility and usability of open access digital library systems in India;
• To examine the use of metadata in the digitization projects for describing digitized contents;
• To study critically the search and retrieval interfaces adopted in the digitization and open access projects in India;
• To study critically the content management mechanisms adopted in the digitization projects; and
• To suggest the effective model for dissemination of indigenous knowledge, documentary heritage and current scholarly literature.

1.5 Scope
This research work aims at producing *State of the Art Report* in the emerging area based on exploratory research method. The objective of exploratory research is to gather preliminary information that will help define problems and suggest hypotheses. Given its fundamental nature, exploratory research often does not require any hypothesis (Kotler & Armstrong, 2006). Thus, this research work was carried out without any hypothesis.

In this thesis, some well-known projects have been evaluated based on some parameters as indicated in Section 1.4 of this Chapter. Apart from some national level initiatives, some institutional cases are also covered in this thesis while these institutional initiatives are strongly representing national ethos on this emerging area.

1.6 Methodology
This research work has been carried out in different phases.

Phase I: In this phase the author was involved in review of literature with special focus on national and international policy instruments, digitization standards, digitization guidelines, best practices, country reports, institutional case studies,
and current research findings related to digital library development in the country and abroad.

Phase II: In this Phase the author prepared a structured questionnaire and tested this questionnaire in consultation with digital library practitioners in India. After successful testing, questionnaires were circulated to project coordinators of selected digitization initiatives across India. This questionnaire is provided the Annex I of this thesis. Later, the author undertook field visits to the important digitization project sites across India for collecting data, applying observation and interview methods. The basis for selecting the institutes for study and field work was (i) representative project sites of national level initiatives such as Digital Library of India’s regional mega scanning centres, scanning centres and source libraries; Indian Institute of Science (IISc), Indira Gandhi National Centre for the Arts (IGNCA), National Mission for Manuscripts (NMM), National Informatics Centre (NIC) and many other important institutions across the country; (ii) prior appointments with the key persons in the national level initiatives – whoever responded in time.

In this field work, the author interviewed project coordinators, project supervisors and project staff – whoever were available – for recording strategic, operational and practical information related to digitization work. In this phase critical and exhaustive studies were carried out mostly following case study method. Case studies revealed detailed descriptions of digital library projects. Each component in a digitization project set up such as ICT infrastructure, digitization equipment, software used for content digitization, quality control and aggregation; information storage and retrieval system are critically observed and recorded. Observations in the field visits were properly documented as per scientific research method.

Phase III: This phase revolves around critical analysis of gathered data. In this phase, final report is prepared in the form of thesis incorporating analytical description of selected digitization projects, recommendations and suggesting future research directions.
In this thesis *APA Style* is followed for citing references and sources of information, which is based on fifth edition of the Publication Manual of the American Psychological Association (APA, 2001).

1.7 Review of Literature

1.7.1 International and National Policy Instruments

The policy instruments are intervention tools used to achieve national and institutional objectives of a given thrust area. The key policy instruments for the open access to knowledge and information are international declarations accepted in intergovernmental forums, recommendations of policy advisors to highest government authorities and government regulations. National level organizations and local institutions may use a variety of policy instruments to achieve their objectives. It is very fortunate that Indian digitization and open access initiatives have been given much importance on various provisions of these policy instruments and adopted appropriate measures in implementing them.

1.7.1.1 International Policy Instruments

a) WSIS Declaration of Principles and Plan of Action

The United Nations-backed World Summit on the Information Society (WSIS) strongly supported open access to information and knowledge (United Nations, 2005). This confirms that member countries of the United Nations will take appropriate strategic decisions to bring scholarly literature, produced from public-funded research initiatives or state-supported researchers, under the umbrella of open access. In this instrument, some specific principles and plan of action were suggested relevant to digital preservation of cultural contents, creation of local contents and open access to scholarly literature. An excerpt of relevant texts from both *Declaration* and *Plan of Action* is highlighted below.
**WSIS Declaration of Principles**

“We strive to promote universal access with equal opportunities for all to scientific knowledge and the creation and dissemination of scientific and technical information, including open access initiatives for scientific publishing.” [B3. Access to information and knowledge]

“The preservation of cultural heritage is a crucial component of identity and self-understanding of individuals that links a community to its past. The Information Society should harness and preserve cultural heritage for the future by all appropriate methods, including digitisation.” [B8. Cultural diversity and identity, linguistic diversity and local content]

**WSIS Plan of Action**

“Support efforts to develop and use ICTs for the preservation of natural and cultural heritage, keeping it accessible as a living part of today’s culture. This includes developing systems for ensuring continued access to archived digital information and multimedia content in digital repositories, and support archives, cultural collections and libraries as the memory of humankind.

Develop and implement policies that preserve, affirm, respect and promote diversity of cultural expression and indigenous knowledge and traditions through the creation of varied information content and the use of different methods, including the digitization of the educational, scientific and cultural heritage.

Support local content development, translation and adaptation, digital archives, and diverse forms of digital and traditional media by local authorities. These activities can also strengthen local and indigenous communities.” [C8. Cultural diversity and identity, linguistic diversity and local content]

“Encourage initiatives to facilitate access, including free and affordable access to open access journals and books, and open archives for scientific information.” [C3. Access to information and knowledge]

---

1 World Summit on the Information Society (WSIS) - Declaration of Principles
<www.itu.int/wsis/docs/geneva/official/dop.html>
2 World Summit on the Information Society (WSIS) - Plan of Action
<www.itu.int/wsis/docs/geneva/official/poa.html>
"Promote electronic publishing, differential pricing and open access initiatives to make scientific information affordable and accessible in all countries on an equitable basis." [C7. E-science]

b) UNESCO Universal Declaration on Cultural Diversity

The United Nations-backed Universal Declaration on Cultural Diversity strongly promotes the concept of digitizing documentary heritage collections for initiating intercultural dialogue which is the best guarantee of peace (UNESCO, 2001). This Declaration confirms that member countries of the United Nations will take appropriate strategic steps to bring cross-cultural documentary heritage collections accessible to worldwide communities. An excerpt of relevant texts from both Declaration and Main Lines of an Action Plan is highlighted below:

“Towards access for all to cultural diversity While ensuring the free flow of ideas by word and image care should be exercised that all cultures can express themselves and make themselves known. Freedom of expression, media pluralism, multilingualism, equal access to art and to scientific and technological knowledge, including in digital form, and the possibility for all cultures to have access to the means of expression and dissemination are the guarantees of cultural diversity.” [Article 6 of the Declaration]

“Encouraging “digital literacy” and ensuring greater mastery of the new information and communication technologies, which should be seen both as educational discipline and as pedagogical tools capable of enhancing the effectiveness of educational services. [Main Line # 9]

Promoting linguistic diversity in cyberspace and encouraging universal access through the global network to all information in the public domain. [Main Line# 10]

Countering the digital divide, in close cooperation in relevant United Nations system organizations, by fostering access by the developing countries to the new technologies, by helping them to master information technologies and by facilitating the digital dissemination of endogenous cultural products and access by those countries to the educational, cultural and scientific digital resources available worldwide.” [Main Line # 11]

1 UNESCO Universal Declaration on Cultural Diversity
<http://unesdoc.unesco.org/images/0012/001271/127160m.pdf>
c) Open Access Statements and Declarations

In the wake of the open access movement, some policy frameworks have already been established by member communities to foster inclusive, plural and development-oriented knowledge societies. A number of open access declarations/statements were made during the past decade, where the world’s leading research institutions agreed on the open access mandates. An indicative list of major open access statements or declarations is given here:

- ARIIC Open Access Statement (Australian Research Information Infrastructure Committee) [www.caul.edu.au/scholcomm/OpenAccessARIICstatement.doc]
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [http://oa.mpg.de/openaccess-berlin/berlindeclaration.html]
- Bethesda Statement on Open Access [www.earlham.edu/~peters/fos/bethesda.htm]
- Budapest Open Access Initiative Statement [www.soros.org/openaccess/]
- ERCIM Statement on Open Access (European Research Consortium for Informatics and Mathematics) [www.ercim.org/publication/Ercim_News/enw64/ercim-oa.html]
- IFLA Statement on Open Access to Scholarly Literature and Research Documentation [www.ifla.org/V/cdoc/open-access04.html]
- OECD Declaration on Access to Research Data from Public Funding (Organisation for Economic Co-operation and Development) [www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html]
- Wellcome Trust Position Statement in support of open and unrestricted access to published research [www.wellcome.ac.uk/doc_WTD002766.html]

d) Hyderabad Declaration, 2004

The ICT Ministers in the Asian Region met in Hyderabad in the Asia IT Ministers’ Second Summit during 12-13 January 2004 and adopted a collective principle of undertaking definitive and firm steps to spread the benefits of ICT. In this
declaration, digitization of culture is one of the bold provisions to spread the benefits of ICT to the common citizens. An excerpt of relevant texts from the Declaration is highlighted below:

**Digitization of Culture**¹

“This Summit recognizes the role of culture – that is, the entire spectrum of its heritage, values, practices, ways of life, knowledge systems, languages and artistic expression transmitted through generations - in promoting self-respect in communities and among nations. We realize the great significance of Asia as a centre of cultural diversity. Taking into consideration the fact that traditional expressions of culture are getting obscured, diluted and transformed, this Summit proposes to develop an Asia Digital Culture Community for developing regional cooperation, coordinating international collaboration and catalyzing knowledge preservation in member countries. The preservation of our collective cultures will be enhanced through such activities as:

- Raising awareness among international communities and groups about the value of digital documentation of their culture.
- Adopting an agreed code of ethics for collection, digital documentation and publication of cultural resources.
- Developing local language technologies and on-line transmission capabilities.
- Fostering training in conservation, digitization and dissemination.
- Enabling IPR development for free access and dissemination of cultural resources.”

### 1.7.1.2 National Policy Instruments

**a) NKC recommendations on Libraries and Open Educational Resources**

The National Knowledge Commission of India (NKC), constituted on 13th June 2005, is a high-level advisory body to the Prime Minister of India, with a mandate to guide policy and direct reforms. NKC’s overarching aim is to transform India into a vibrant knowledge-based society. The NKC’s Working Group on Libraries

---

and Working Group on Open Access and Open Educational Resources have strongly recommended digitization of documentary heritage collections, open access to public-funded research literature and supported establishment of open courseware repositories for countrywide dissemination of quality courseware to many cross sections of people. If implemented, these recommendations will have far-reaching implications in the knowledge creation and dissemination cycle. The documentary heritage collections, scholarly literature and lifelong learning materials digitized or produced by state-sponsored institutions would then be made accessible through open access channels such as national and institutional repositories. This way the NKC’s recommendation on peer-reviewed research papers resulting from public-funded research would be validated by subject experts when making these resources available through open access channels. NKC also recommended creation of national knowledge portals for basic needs/ key sectors such as water, energy, environment, education, food, health, agriculture, employment and citizen rights. Already national portals on water (India Water Portal, www.indiawaterportal.org), energy (India Energy Portal, www.indiaenergyportal.org), environment (India Environment Portal www.indiaenvironmentportal.org.in) and education (Sakshat, www.sakshat.ac.in) have been launched that provide open access to information, knowledge and learning resources on the relevant areas. An excerpt of relevant recommendations and open access statements of NKC (2006, 2007) is highlighted below:

**Digitization and Open Access**

“To enable equitable and universal access to knowledge resources, it is important to create more digital resources which can be shared. The concept of an ‘information commons’ i.e. ‘resources shared by a community of producers and consumers in an open access environment’ needs to be promoted. New resources should be openly accessible and historical documents, too, should be digitized and made available.

---


License: CC BY-NC-ND
• This Working Group strongly recommends that peer-reviewed published research papers resulting from publicly funded research in India must be made available through open access channels, subject to copyright regulations. The group also recommends use of open standards and free and open source software.

• All pre-independence periodicals and newspapers in all Indian languages and in English must be digitized for access and preservation.

To help preserve digital resources, optimize their use and avoid duplication of effort:

• State-level archives for preservation of digitized materials must be set up.

• Every state should establish a registry and archives of knowledge-based digital resources, and make it accessible.”

Encourage Open Access

“Open access material stimulates research and helps students, teachers and researchers across the world. Therefore at a policy level, all research articles published by Indian authors receiving substantial government or public funding must be made available under Open Access and should be archived in the standard OA format at least on his/her website. As a next step, a national academic OA portal should be developed. The government should allocate resources to increase the current digitization efforts of books and periodicals which are outside copyright protection. Separate funding should be allocated to develop a new high quality OCR software package so that new and old fonts in many different Indian languages can be converted into ISCI/ASCI code and OA portals and servers could be upgraded regularly. Appropriate financial resources should be earmarked for these endeavours. This will also facilitate machine translation of these valuable resources.”

Summary of NKC Recommendations to Prime Minister of India

On Libaries (2006)

• Set up a National Commission on Libraries

• Prepare a national census of all libraries

---


License: CC BY-NC-ND
- Revamp library information science (LIS) education, training and research facilities
- Re-assess staffing of libraries
- Set up a central library fund
- Modernize library management, encourage greater community participation in library management
- Promote information and communication technology (ICT) applications in all libraries
- Facilitate donation and maintenance of private collections
- Encourage public private partnerships in LIS development

**On Open Educational Resources (2007)**
- Support the production of quality content by a select set of Indian institutions
- Leverage global open educational resources
- Encourage open access
- Develop network-enabled delivery infrastructure
- Create a faculty and institutional development programme

**On Other Thrust Areas – having relevance on Open Access**
- Protection of traditional knowledge (TK) through Traditional Knowledge Digital Library (TKDL) and promoting incentives for wealth creation from TK [Intellectual Property Rights, 2007]
- Set up a National Education Foundation to develop web-based common open resources [Open and Distance Education, 2007]

b) UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005

In 2005, the University Grants Commission of India (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 such as:

- Creation of Indian National Theses Database (INTED)
- Submission of PhD Theses in Electronic Form

This set of regulations is still under consideration. They propose to formulate a roadmap to achieve widest dissemination of results of doctoral research conducted in Indian universities and bibliographic control of theses and dissertations of research degrees. So far, a few UGC-supported universities have established open access repositories for scholarly literature produced in the respective universities. Initially the universities were reluctant to change their status quo, as the contents of scholarly literature including the PhD theses will be a matter of critical analysis by the national and international peers, if the universities establish open access repositories. Now, universities will be motivated to produce qualitative doctoral theses and will maintain certain international standards. The UGC also supports development of infrastructure in Indian universities through various planned schemes. The proposed national education grid will also enrich modern ICT infrastructure in Indian universities. The Indian universities then will have necessary infrastructure to host a number of Web-based information services. Hosting an ETD repository and providing online interface to INTED will not be a problem in most of the universities. Indian universities should now take a proactive role in the implementation of these regulations at the earliest, ensure qualitative research and make the results of doctoral research widely available. The UGC should also clear all bottlenecks to implement INTED and decentralized ETD repositories across the country.

1.7.2 Digitization Guidelines and Best Practices

The international communities have developed several digitization guidelines and best practices to guide their member institutions on the effective ways to implement digitization projects. Some guidelines deal with technical specifications and other operational issues in great details, whereas some other
guidelines help the project managers in handling other aspects of the digitization projects such as human resources, finance, legal matters including intellectual property rights, advocacy, and how to involve stakeholders.

### 1.7.2.1 International Guidelines

Several guidelines are in existence at the national and international levels. Some international guidelines are followed in Indian digitization projects, whereas some national guidelines have adopted international guidelines with some degree of localization. Popular international guidelines are described in this Section.

**a) Guidelines for Digitization Projects for Collections and Holdings in the Public Domain, particularly those held by Libraries and Archives**

This document is jointly prepared and published by UNESCO, IFLA (International Federation of Library Associations and Institutions) and ICA (International Council on Archives) in March 2002. These Guidelines are the result of a joint venture of a group of experts on behalf of IFLA and ICA, who had been invited to draft these for UNESCO (2002). This document contains eight comprehensive chapters such as:

- Selection of documents for digitization
- Technical requirements & implementation
- Legal aspects
- Budgeting
- Human resource planning
- Development & maintenance of Web interfaces
- Preservation of digital content
- Project management.

**b) The Guide for Electronic Theses and Dissertations**

This guidebook is the result of a joint venture of a group of experts on behalf of UNESCO and NDLTD (Networked Digital Library of Theses and Dissertations). This document is prepared for ETD stakeholders such as students, researchers research supervisors, policymakers, ETD practitioners and NDLTD members (UNESCO, 2002). This guidebook provides a roadmap for implementing ETD at
institutional and national scale. Targeting different stakeholders, this guide contains several comprehensive chapters such as:

- Universities
- Students
- Technical issues
- Training the trainers
- The Future

c) SPARC Institutional Repository Checklist & Resource Guide

This is a comprehensive guidebook prepared by Raym Crow (2002) for the member institutions of the Scholarly Publishing & Academic Resources Coalition (SPARC). This guidebook helps the worldwide communities including Indian institutions in implementation of institutional repositories as well as open access subject repositories for wider dissemination of scholarly literature. The chapters in this document include:

- Securing Administration Support
- Securing Faculty Participation
- Librarians: Benefits and Challenges
- Repository Management and Policy Issues
- Technical and System Issues

1.7.2.2 National Guidelines

National level projects in India also prepared and published some guidelines based on international guidelines and best practices to achieve uniformity in digitization work carried out simultaneously at different digitization centres. Some important guidelines emanated from Indian digitization projects are:

a) Million Book Universal Library Project: Manual for Metadata Capture, Digitization, and OCR

This Manual was prepared by Gabrielle V. Michalek (2003) for the Digital Library of India and Universal Digital Library projects. This manual is used across scanning centres and regional mega scanning centres of DLI project in India.
This Manual covers operational aspects of mass digitization work for rare books and printed documents. This document contains five chapters, namely:

- Data Production
- Getting MARC Records from OCLC
- Creating Metadata Using Dublin Core
- Minolta PS 7000 – Quickscan Software Instructions
- ABBYY FineReader 6.0 Instructions

b) Guidelines for Digitization of Manuscripts

This document was prepared by the digitization team of National Mission for Manuscripts (2004). This manual is used by manuscript digitization centres of the Mission across India. This is a very comprehensive manual covering topics namely:

- Criteria for Selection
- Technical Requirements and Implementation
- Output Specification
- Metadata Creation
- Collection Management
- Image Quality Check List

c) UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005

This document was prepared in 2005 for national level consultation of stakeholders in Indian higher education system. This document provides a detailed roadmap for achieving ETD repositories at the national and university levels. This document is divided into two parts. The first part provides background information on Indian and international scenarios. In this part, some specific standards for theses data are also elaborated. The second part deals with implementation process. An outline of this manual is given below:

- Part I: Background Information, Current Scenario, Major Issues & Data Standards
  - Present Scenario in India
1.7.3 Literature Describing National Digitization Initiatives

Indian digital library and digitization initiatives are widely described and evaluated by the national and international researchers in their papers. Most of the projects described in the literature are successfully implemented national level initiatives (Vikas, 2005). Open access initiatives are also highlighted in some of the papers (Fernandez, 2006; Ghosh & Das, 2007; Rajasekhar, 2003; Suber & Arunachalam, 2005; Ahmed & Rather, 2007), describing different kinds of projects such as institutional repositories, subject repositories, open educational resources, open access journals, metadata harvesting services and library consortiums. Some of the literature communicated in journals and conferences can be classified as country report holistically describing country efforts in the areas of digital library development and open access movement. The several directories also provide comprehensive listing of open access projects implemented in India adhering to international standards, more particularly Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH). Examples of such directories are Directory of Open Access Journals (DOAJ), Directory of Open Access Repositories (Open DOAR), Registry of Open Access Repositories (ROAR), and Cross Archive Search Services for Indian Repositories (CASSIR). These directories have become effective information sources for several papers (Fernandez, 2006; Ghosh & Das, 2007; Mittal & Mahesh, 2008) describing country efforts presented in conferences and published in Indian journals.
Indian government agencies supporting digitization and open access projects across the country also provide comprehensive listings of completed and ongoing projects supported by them. For example, National Digital Libraries Cell of the Department of Information Technology in MCIT publishes a list of projects funded by MCIT and their current status including achievements\(^1\). Among the listed projects funded by MCIT, most prominent ones are Digital Library of India (DLI) and its associated projects, cultural heritage digital library projects at IGNCA. In this thesis both the initiatives has been dealt in great details.

Another agency supporting digitization and open access projects is the Technology Information Facilitation Programme (TIFP) under Department of Scientific & Industrial Research (DSIR), Government of India. This Programme is the successor of the National Information System for Science and Technology (NISSAT) programme of DSIR. TIFP publishes a list of digital information management projects funded by DSIR and their current status including achievements\(^2\). DSIR is also country focal point for the Asia Pacific Information Network (APIN), promoted by UNESCO’s Information for All Programme (IFAP). Tuning with APIN’s long-term strategies, TIFP’s major thrust areas, emphasizing on content creation, digitization, digital library and open access development in the country, include:

- Indian digital library of theses and R&D publications;
- Electronic publishing of selected Indian S&T materials;
- Open archive initiatives – a Web alternative to scholarly communications;
- Documentation of traditional knowledge and folk wisdom;
- Promotion of information access and sharing;
- Promotion of content development;
- National websites/ servers.

Some of the successful digitization projects funded by DSIR and analyzed in this thesis are:

- Development of OAI-Based Institutional Research Repository Services in India, undertaken by National Centre for Science Information (NCSI), Indian Institute of Science.
- Vidyanidhi Digital Library, undertaken by University of Mysore.

1.7.3.1 International Conferences held in India dealing with Digital Libraries

Conference papers are major source of information on digital library initiatives in India. Many international conferences, several national conferences as well as national seminars held in India have produced a huge amount of literature greatly dealing with theoretical aspects of digitization, digital library development and open access archives. Only a limited portion of the conference literature deals with research problems as well as actually implemented cases. Few qualitative research papers presented in these conferences later got published in national and international peer-reviewed journals, whereas some other conference papers were archived in subject repositories such as Librarian’s Digital Library and E-LIS. A list of international conferences on digital libraries held in India is given here:

- The Fourth International Conference of Asian Digital Libraries (ICADL 2001) was held in Bangalore during 10-12 December 2001. This conference was organized by University of Mysore and Indian Institute of Information Technology (IIIT) Bangalore. The theme of the Conference was Digital Libraries: Dynamic Landscape for Knowledge Creation, Dissemination and Management. ICADL is a series of international conferences with an Asian focus, first one held in Hong Kong in 1998. ICADL 2001 was the first major international conference held in India on digital libraries. In this conference, several Indian professionals presented papers mostly in the areas of theoretical understanding on the subject (Urs, Rajashekar & Raghavan, 2001).
• The First International Conference of Digital Libraries (ICDL 2004) was held in New Delhi during 24-27 February 2004. This conference was organized by TERI (The Energy and Resources Institute), New Delhi. The theme of the Conference was knowledge creation, preservation, access and management. This was the one of the major international conferences held in India on digital libraries. In this conference, over one hundred Indian professionals presented papers mostly in the areas of theoretical understanding on the subject. In some papers institutional digital library and digitization initiatives are also highlighted. The Conference proposed a National Mission on Digital Libraries in coalition with the other Missions initiated by the Department of Culture (Kar, 2004).

• The Second International Conference of Digital Libraries (ICDL 2006) was held in New Delhi during 5-8 December 2006. This conference was organized by TERI, New Delhi. The theme of the Conference was information management for global access. In this conference, over one hundred Indian professionals shared their practical experiences in the implementation of institutional digitization and institutional repository projects. The Conference proposed the enactment of a Digital Library Act to facilitate digitization and content creation for universal access to information (Kar, 2006).

• The First IEEE International Conference on Digital Information Management (ICDIM 2006), organized by the Digital Information Research Foundation Chennai and Chirst College Bangalore, was held in Bangalore during 6-8 December 2006. One of the themes of the Conference was digital libraries. In this conference, a number of professionals presented cases on Indian digital libraries and digital archives.

• The International Conference on Information Management in a Knowledge Society (ICIM 2005), organized by Indian Association of Special Libraries and Information Centres (IASLIC) to commemorate its golden jubilee celebrations, was held in Mumbai during 21-25 February 2005. One of the
themes of the Conference was digital libraries and digital archives. In this conference, a number of Indian professionals presented institutional cases on digital libraries and open access archives (Chandra, 2006).

- The International Conference on Semantic Web and Digital Libraries (ICSD 2007), organized by Documentation Research and Training Centre (DRTC), Indian Statistical Institute, was held in Bangalore during 21-23 February 2007. This Conference had two broad themes, viz., semantic Web and digital libraries and the convergence of both the technologies. In this conference, a number of Indian professionals presented theoretical frameworks on semantic Web concept and its integration with digital libraries and open access archives. Case studies of a few semantic Web applications in Indian open access and digital library initiatives were also presented in this Conference.

- The International Conference on Building Knowledge Repositories: Cross-Sectoral Collaborations, organized by National Institute of Fashion Technology (NIFT), was held in New Delhi during 7-9 February 2008. In this conference, a number of Indian professionals presented institutional cases on knowledge repositories and open access archives as well as shared their experiences on the cross-sectoral partnership in knowledge dissemination. The Conference recommended (i) facilitating and enhancing cross-sectoral collaborations among institutions for developing digital repositories through a nodal agency; (ii) establishing a national depository for digital objects being produced by scholarly societies and institutions; and (iii) promoting open access standards for convergence and collaboration through research and training.

1.7.3.2 Baseline Surveys and Country Reports on Digitization and Open Access Initiatives in India

A number of Indian researchers have published papers on state of the affairs of digitization and digital library initiatives in India. Some of the authors
contextualize the situations with their organizations they are working in their papers.

First beads of papers on the digitization issues were written by veteran library professionals citing their concerns on the implementation processes of this new domain. Dasgupta (2001, 2004) in her papers raises some concerns on the sustainability issue. To her viewpoint, sustainability is always in doubt as most of the digitization projects are started by getting a one time grant without a follow up plan. She points out that sustainability is a broad term which refers to many factors starting from technological issues of preserving digital data to social and economic questions for long-term accessibility of information to the people. For tackling sustainability factors, she suggests a focused attention on the broad spectrum comprising:

a) Policy and strategies,
b) Technical issues,
c) Economic aspects,
d) Social and educational issues,
e) Political and administrative will,
f) Capacity building,
g) Cooperative ventures, and
h) Changing role of librarians and information professionals.

Research Programme, Vidyanidhi Digital Library, Muktabodha Digital Library and Archiving Project, and e-Journals@INSA. Mittal and Mahesh (2008) also published literature review pertaining to digital library development in India, in addition to descriptive accounts of digital library initiatives in the country.

On the other hand, open access to information and knowledge is related concept to digitization although situated in different ideal plane. While digitization of library materials aims at long-term preservation of rare documents that are on the verge of extinction, open access to knowledge aims at providing access to current literature mostly available as born digital objects.

Mittal and Mahesh (2008), Ghosh and Das (2007), Ahmed and Rather (2007), Fernandez (2006), and Das et al. (2005) published baseline surveys on open access initiatives in India, indicating digital archiving software used for establishing open access archives. Ghosh and Das (2007) also analyzed reasons for India’s leadership in open access movement and the present research situations in India in the changing global context. Fernandez (2006) received research grant from the Shastri Indo-Canadian Institute for the evaluative study of open access initiatives in India and he did some in-depth study on the subject emphasizing on existing policy frameworks versus policy requirements.

Arunachalam (2008, 2005) in his writings emphasizes on necessities of open access to scientific knowledge in building research capacity in developing countries including India. He also campaigns for the implementation of open access projects at the regional, national and institutional level. He publishes several articles on this topic in different journals and forums. Das (2008) publishes a book on open access to knowledge and information with a comprehensive directory of open access and digital library initiatives in South Asia. This book has comprehensive coverage on different kinds of initiatives such as digital library projects, open courseware, open access journals, metadata harvesting services, national level open access repositories and institutional repositories.
1.7.3.3 Institutional Case Studies on Digitization and Open Access Initiatives in India

In 2005, *Journal of Zhejiang University Science* published special proceedings issue of the first International Conference on Universal Digital Library (ICUDL 2005). In this issue several papers described research carried out in Digital Library of India project. Balakrishnan (2005) describes technological challenges and future research directions. Other papers describing research problems include: self healing information system for digital libraries (Prahllad and Black, 2005), text-to-speech (TTS) system for universal digital library (Ambati and Reddy, 2005), optical character recognition (OCR) system for Tamil text (Seethalakshmi et al., 2005), Om transliteration tool for Indian languages (Ganapathiraju et al., 2005), transliteration editors for Indian languages (Prahllad, 2005). *Vishwabharat@TDIL*, a journal of Technology Development for Indian Languages Programme also devoted a special issue on fonts technology and digital library of India. Coordinators of regional mega scanning centres in Digital Library of India project shared their project experience in this issue. Digital Library of India project is critically analyzed and described in Chapter two of this thesis.

Many other researchers have published institutional case studies describing their projects in different phases, from proposal stage to implementation phase. Arora (2003) provides a detailed account of South Asia’s largest library consortia ‘Indian National Digital Library in Engineering Science and Technology’ (INDEST) from the initial proposal stage to implementation road maps. This paper also describes digital library and digitization activities carried out in Indian Institute of Technology Delhi, funded by different government agencies. Mujoo-Munshi (2003) provides a detailed account of journals digitization project of Indian National Science Academy (INSA) titled e-Journals@INSA. This paper describes other digitization activities proposed and carried out in INSA, funded by different government agencies.

Indian Institute of Sciences (IISc), Bangalore is one of the pioneering institutions in India having successful open access initiatives. IISc open access initiatives
have become standard models for many institutional repositories in India and even aboard. Jayakanth (2008), Anuradha (2005), and Rajasekhar (2003) have published papers describing open access publishing model adopted in IISc for worldwide dissemination of research papers published by researchers and faculty members of IISc. They also described operational aspects of different open access initiatives in IISc such as EPrints@IISc, ETD@IISc and Pravabhi: IISc Publications database.

In two separate papers, Deb (2006) and Deb and Kar (2005) describe digitization activities and integration of digitized contents into existing hybrid digital library at The Energy and Resources Institute (TERI) in New Delhi.

In his paper, Sutradhar (2006) shares his experience on implementation of a intranet-base institutional repository in IIT Kharagpur using DSpace open source software.

In her paper, Patra (2006) shares her views on need for a digital library on ceramics. She also proposes a roadmap for implementing a digital library in ceramics in her institute.

National Library of India (NLI) is an institution of national importance under Ministry of Culture having rarest of collections of manuscripts and books. In 1999 NLI undertook a pilot project entitled “Down Memory Lane” to digitize its rare and brittle books. From February 1999 to June 2001, a total of 6601 books containing more than 2.5 million pages were scanned and archived in 548 CD-ROMs (in duplicate). This digitized collection can be consulted in NLI premises only. In a paper (National Library of India, 2003), NLI describes its digitization activities for manuscripts and rare books along with some technical details about the project.

There are many other institutional case studies mostly communicated in conferences and national journals (Doctor & Ramachandran, 2008; Bansode, 2008; Doctor, 2008; Doctor, 2007; etc.). In this section, only some significant cases are reviewed that have linkages with this research work.
1.8 Chapterization

This thesis addresses essentially two issues – the digitization of library materials available with memory institutions and digitization of scholarly materials produced by Indian researchers and making these scholarly materials available through open access channels. This section presents an overview as to how the various parts and sections of this research are related to the research problems and observations of the evidences leading to the recommendation of functional as well as business models for self-sustainability of the digitization projects. As indicated in Figure 1.1, each chapter, within the chapters 2 and 5, presents the evidence from perspectives of implemented projects. Chapter 2 presents critical appraisal of India’s largest digitization initiative – Digital Library of India. Chapter 3 presents critical appraisal of digitization work pertaining to theses and dissertations – that makes the way to the creation of electronic theses and dissertations (ETD). In this chapter Indian initiatives such as Vidyanidhi Digital Library, ETD@IISc, OpenMed@NIC, and CSIR e-Thesis projects are described.

Chapter 4 presents critical appraisal of digitization work of documentary heritage collections. This chapter deals with digitization work for most venerable pieces of artifacts such as rare manuscripts, rare books, old photographs, photographic slides, audio tapes and microfilms. In this chapter, digitization initiatives of two prominent Indian institutions, namely, Indira Gandhi National Centre for the Arts (IGNCA) and National Mission for Manuscripts (NMM) have been evaluated.

Chapter 5 presents critical appraisal of digitization work of Indian periodicals, more specifically the projects carried out by the Indian National Science Academy (INSA), Indian Academy of Sciences (IAS), Indian Mediars Centre at the National Informatics Centre (medIND@NIC), Medknow Publications and Ramakrishna Math. Digitized journal collections make the way to the creation of open access journals in India. Thus, the genesis of open access journal development in India is also indicated in this chapter.

Chapter 6 concludes with the summary of findings and recommendation of futuristic functional models suitable for changing contexts in the country.
Fig. 1.1: Research Design

- **Chapter 1**
  - Introducing the context of the study
  - Disclosing our preconceptions

- **Chapter 2**
  - Problem Area: Rare Books
  - Case Study: Digital Library of India

- **Chapter 3**
  - Problem Area: Theses and dissertations
  - Case Studies: ETD Initiatives

- **Chapter 4**
  - Problem Area: Documentary Heritages
  - Case Studies: IGNCA and NMM Initiatives

- **Chapter 5**
  - Problem Area: Indian Journals
  - Case Study: Open Access to Journals

- **Chapter 6**
  - Summary of Findings, Conclusion and Recommendations
1.9 References


License: CC BY-NC-ND


License: CC BY-NC-ND


License: CC BY-NC-ND


