

# March towards digitization of information resources in India: issues and initiatives

R K BHATT

Head and Reader, Department of Library and Information Science  
University of Delhi, Delhi - 110 007

World Digital Libraries 1(2): 147-164

---

## **Abstract**

In the past two decades ICT (information and communication technology) has completely transformed the library and information science in India and paved the way for the establishment of the digital libraries. The present article attempts to define the concept of digital library and elaborates the functions, opportunities and challenges faced by the digital libraries. In addition to the several issues involved in the process of digitization of the information resources, also discusses the digital library initiatives in India taken by both private and government agencies.

## Introduction

Recent developments in information and communication technologies, especially the Internet and the web, have brought significant changes in the ways we generate, distribute, access and use information. For centuries we have been using printed material made available to us by the systematic efforts of the publishers, book sellers, and librarians. But with the advent of information technology, the situation began to change as the printed information was started to be digitized and made available to use with the help of computer networks. This has changed the whole phenomenon of library and information science.

## Definition of the digital library

In a digital library, librarians provide information to the users in the digital format by exploiting the facilities of information technology. According to 'Association of Research Libraries', a digital library has the following common elements.

- The digital library is not a single entity.
- The digital library requires networking technology to connect many entities.
- All linkages are transparent to end users.
- Universal access to digital content and information is a goal.
- Digital library collections are not limited to document surrogates, but can include digital artefacts not extant in traditional formats.

Several definitions of digital libraries are available in the literature. Many of these definitions were formulated in the course of digital library research projects. Consequently they have been influenced by the people involved in the projects, their understanding of the concept of libraries vis-à-vis electronic databases and also by the nature of the research project. Borgman (2000) analyses a number of definitions of digital libraries and concludes that there are two major classes.

- Those coming from digital library researchers (in the US context they are generally computer scientists and engineers); and
- Those coming from library and information professionals.

Borgman (2000) further opined that the first research-oriented definition of digital libraries came in 1992 when the phrase 'electronic libraries' was used. The phrase 'electronic libraries' is now used as 'digital libraries'. In digital libraries, according to these researchers, focus is given to the contents collected whereas, librarians focused on digital libraries institutions or services. Marchionini and Fox (1999) while strengthening the service point of view of the digital libraries said that digital library work occurs in the context of a complex design space shaped by four dimensions: community, technology, services, and content.

- Marchionini and Fox (1999) suggests that
- The community dimension of digital libraries reflects social, political, legal and cultural issues.
  - Technology serves as the engine moving the digital library field, including technical progress in computing, networking, and more specifically information storage and retrieval, multimedia, interface design, and so on.
  - Services form the central focus of digital libraries and future digital libraries should facilitate digital reference services, real-time question answering, on demand help, information literacy and user involvement mechanisms.
  - Content represents all possible kinds of form and generate information both printed as well as digital.

Oppenheim and Smithson (1999) lays emphasis on digital technologies and said that a 'digital library is an information service in

which all the information resources are available in computer processable form and the functions of acquisition, storage, preservation, retrieval, access and display are carried out through the use of digital technologies'. Gladney, Fox, Ahmed, et al. (1994) gave the most comprehensive definition of the digital library when he said that 'a digital library is an assemblage of digital computing, storage and communications machinery together with the content and software needed to reproduce, emulate and extend the services provided by conventional libraries based on paper and other material means of collecting, cataloguing, finding and disseminating information. A full service digital library must accomplish all essential services of traditional libraries and also exploit the well known advantages of digital storage, searching, and communication'.

In fact, digital libraries, like traditional ones, will select, acquire classify and catalogue and disseminate information and knowledge. The major difference will be that a digital library will consist of machine-readable data. This implies that the traditional concept of a collection must be revised to accommodate materials that are accessible electronically.

### Digitization

A tremendous quantity of information in various formats is created daily through various media and it is becoming increasingly difficult to remain oriented in this flood of information without the help of computer technology. In fact the issue of effective consumption of information is presently more relevant than its longevity and has not been addressed comprehensively so far. As commonsense indicates, the more the information created, the less assured is its longevity. This dilemma necessitates a process of selection; some information must be prioritized and its existence ensured, even

while other kinds of information are sidelined.

Ashman (2003) 'Digitization is the process of converting analogue information to a digital format for storage and processing in a computer'. On the other hand, according to the definition formulated by Pradhan, (2004) opined that 'Digitization is a process that involves a lot of initial and long-term costs for the further maintenance of digital objects'.

### Need for digitization

The basic purpose of digitization is to exploit the ICT facilities for sharing global resources. Libraries are digitizing materials because they remain convinced of the continuing value of such resources for learning, teaching, research, scholarship, documentation and public accountability. Digitization raises the profile of the institutions as world wide users can know the institutional collection and utilize these resources from remote locations, which otherwise was not known or possible earlier with print publications. These merits have led to the significant growth of various national and international digitization projects in the past ten years. Digitize object may reveal information not visible in originals as image manipulation software can improve image tone from light to dark to light, adding more clarity to the image and making the information usable.

- To preserve the documents in order to read older or unique documents without damaging the originals.
- To make the documents more accessible to serve existing users better in other words, to allow users to search the full text of the documents even from remote places by many. Simultaneously to bring together scattered materials on a specific topic.
- To reuse the documents and convert them into different formats, for example, to use images in a slideshow. Like wise adapt the content for a different purpose, for

example, to convert the text of a report into training materials.

### What to digitize

Most of the libraries are having some unique materials, manuscripts, historic documents, personal documents and the collection - fragile, brittle, mutilated and of non-serviceable nature. Such materials need due attention for their preservation and making them serviceable. Digitization is an answer to such type of materials and problems.

However, it is easy to digitize white, clean opaque paper whereas coloured, damaged or thin paper is difficult to digitize. Simple, layout, single columns, single sheets, straight text with headings, few pictures, standard computer typefaces, typewriting, handwriting, and so on, are easy to digitize whereas, complex, layout, multiple columns, fragile, heavy bindings, pictures, equations, tables, unusual typefaces, and poor quality printing are difficult to digitize.

### Requirements

- Equipment – Scanners, computers and storage devices, Audio and video capture equipment.
- Software – Scanning, optical character recognition, word processing, spell check, image management, video and audio capture.
- Human resources – Personnel and skills.
- Funds – To cover salaries, equipment, software, running costs, and so on.

### Steps for digitization

The various steps involved in digitization are:

- Selection of material which needs to be done on the basis of demand, value and condition of the information resources to be selected for the purpose of digitization. The selected material should be either on the public domain or is out of the copyright implications or the prior

permission from the concerned publisher be obtained before digitizing such a information resource.

- Disbinding the material – Once the material is selected, the next step is to decide whether the material require to be disbinded or not. The binding is removed in those cases where the source material is not fragile as it may destroy the original material. To support effective scanning, the spine is to be removed and the pages are precisely and evenly cut. Disbinding makes the physical handling of pages easier but this destroys the source material and therefore, should be done very carefully (2003).
- Scanning – Scanning is the process in which a printed material (analog) is converted into a digital image. This scanned image is nothing more than a photograph that cannot be manipulated. For scanning, scanner is used which comes along with scanning software or capturing software. Generally for printed material like books, journals, and so on, book scanners are available and for disbinded material, that is, loose sheets, an ADE (automatic document feeder) scanner is available in market. Zeutschel OS 5000 and Minolta PS 7000 are examples of book scanners. Once the material is scanned, it is saved into desired format. Most common format used is TIFF (tagged image file format) though it occupies more space.
- Image processing – The scanned image is exactly similar to a photocopy of original document, therefore, if the original document contains unwanted spots, lines, or scratches, the same (dust, lines, and the like) is seen in the scanned image also. Thus, to make it more legible, it needs to be processed through two-step process. First is cropping, where the text is selected leaving behind the noise in margins of paper. The second step is cleaning of the cropped

image. In this step, the dots, strains, spots, or unwanted lines in between the text are cleaned up so that it can be read easily. For cropping and cleaning, one can use two different software or single software (2 in 1) can be used for both the purposes. For cropping, one can use Image Cropper (software) and for cleaning, Scanfix can be used. Adobe Photoshop can be used for both the purposes. After processing, the image is again saved in desired format. The processed image can be viewed by using Irfanview or ACDSee software.

- Optical character recognition – It is a process where the processed image is converted into searchable text. The processed image cannot be linguistically manipulated and therefore it needs to be processed through OCR (Optical character recognition). OCR converts each optically scanned character into an electric signal and compares with internally stored representations of alphabets. The next step is manual checking or proofing of output. This is similar to spelling and grammar option of MS-Word. The software will not only show mistakes by highlighting it in different colour (from the text) but will also show it along with the original text, so that one can select or reject it. This part of OCR process is time consuming and demands more labour. Besides making the text searchable, the OCR also reduces the file space to make it easy to be delivered over Internet. For example, a scanned image of 220 pages stored in TIFF format occupies 12 MB space and the same after processing gets reduced in size to 8.38 MB (in TIFF format). Most commonly used software for OCR is Abbey Finereader.
- Metadata creation – Metadata means data about data. It is similar to the library catalogue that contains the bibliographic details of the material. Metadata is in electronic form and is written in either HTML (hyper text mark-up language) or XML (extensible mark-up language) to specify the structure of individual documents and control how they look like when presented to user. HTML, as compared to XML, is mainly concerned with the physical layout of document like font size, colour, letters in bold, italics, underline, and so on, whereas XML provides a flexible framework for describing document structure and metadata, making it suitable to digital libraries. Another important feature of XML is that it supports Unicode, thus XML can be written in almost any language. The purpose of using XML or HTML is to bind together images and give access to structural elements in single objects or in whole collection (Deegan and Tanner 2002). Metadata, for the easy retrieval of documents is prepared either in MARC (Machine Readable Catalogue) or DC (Dublin Core). The main difference between MARC and DC is MARC is tag based whereas DC has no tag numbers. Second, the number of fields in MARC is much higher whereas in DC, these are only 15, all can be repeated and none is mandatory.
- Output form – Once the files are created, they are stored either in the form of CD, DVD, tape or hard disk depending upon the requirement of libraries.
- Managing storage and delivery systems – Once the digitization process is over, the institution may deliver the images either to Intranet (for local use) or Internet (for global use). Once it is done, it becomes really important to manage and maintain the system for future. Generally, the institutions have AMC (annual maintenance contract) to maintain and update the software, by paying license fee, either with the outsourced agency or with the software (UNESCO, IFLA, and ICA, 2002).

### Issues related to digitization of information resources

There are several issues related to digitization of information resources which need to be taken into consideration before the decision of digitization of the information resources is to be taken.

#### *Shifting from traditional to digital medium*

It is essential that there should be a perfect partnership with the computing service. The partnership can take a variety of forms ranging from organizational mergers to strategic alliances. Conversion of traditional libraries to digital libraries is possible but it is very difficult especially in India, where the literate population has access to more than 71 069 libraries in the country, comprising: 8267 academic, 54 845 public, 1200 science and technology, 450 social science, 800 government departmental, 500 art/culture/humanities, seven national and 5000 industrial and private institutional libraries. However, these cannot be called traditional libraries in the real sense, for more than 90% of them are simply reading rooms. Whereas in the US the total number of libraries are 122 576 (American Library Association) and provide the same access to information, computer technology and services to all library users.

Digital libraries depend on Internet and Intranet connections, yet a full proof system to prevent virus damage is essential and high speed and uninterrupted communication system is also required to provide continuous services.

#### *Maintenance*

In a traditional library there are many development practices and ideals such as circulation, technical service and even shelving to maintain access to the collection

over a period of time. So there is considerable control over the collection. In this way a traditional library could never cope with the limitations of traditional practice and would require too many resources. The digital library may need to find new maintenance mechanisms. Traditional libraries are called narrowly construed libraries, where the collection has known boundaries and possibility of control over the collection is easy. In the broadly construed digital library, users are able to access diverse material, which lead to serious control and long run maintenance issues. This also leads to information overflow.

#### *Universal access to knowledge*

The main development issue is the object of providing universal access, in which libraries play a crucial role. India has launched an ambitious plan to achieve this within 10 to 15 years. It is proposed in the Ninth Five Year Plan report to provide Internet access to every school, university and hospitals. It is being made a compulsory subject to all courses of undergraduate level courses and networking facilities are being extended to most of the centres of higher education. There are at present 877 309 schools comprising primary, middle, and higher secondary schools with an aggregate enrolment of about 80 lakh students. There are about 226 universities, 6569 general colleges and 1354 professional colleges. In addition, there are 835 teacher-training colleges and about 140 000 hospitals. Bringing Internet access to all is a daunting task. This problem is not limited to India but other developing countries too are facing the same. For the user of the digital library, it is essential to have a personal computer with an internet connection. According to the Action Plan on IT, India is likely to have about five million Internet users by the end of year 2008 (Planning Commission, 2008).



### ***Language barrier***

‘The limits of my language mean the limits of my world’ said Ludwig Wittgenstein, (Ludwig Wittgenstein) so all the written material in other languages is inaccessible to use. There are about 6000 languages spoken in the world. However, it is estimated that just over half of the world’s population speak one of just five languages: Chinese, English, Hindi, Russian and Spanish.

In India there are about 18 official languages and more than 1000 dialects spread all over the country. Is it possible to provide material through digital libraries in all languages? Unless this is achieved, digital libraries and universal knowledge will be restricted to the more affluent class familiar with one or more of the major languages of the world. The challenge is to evolve a strategy so as to provide information in the language required by the clientele.

### ***Private participation***

Libraries in the past came into being at the initiative and interest of Kings. With the change to democracy, the responsibility fell on the government. The corporate and private sector was not very involved with the development of libraries, because it was not the top priority at that time. Now the private sector particularly in advanced countries manufacturing Internet related equipment is in a position to extend both materialistic and financial support. Being the major producers of computer technology they are able to set all industries in developing countries. Developing countries could offer some incentives as tax breaks, investment subsidies and so on to these companies and the private and corporate sector could then come forward to make it a reality.

Within about a year after the introduction of the Action Plan for Knowledge for all by the Government of India, most of the public telephone centres in the metropolitan cities in

India, had Internet facilities. However, it is yet to reach rural areas. In this connection the private software sector has come out with some developments in two regional languages, that is, Kannada and Telugu. Also, it could facilitate the involvement of the private sector in promoting and operation of digital library services.

### ***Self-sustained growth***

Information is saleable in today’s materialistic world and in such dynamic environment libraries are playing pivotal role in quenching the quests of its users. In developing countries, the libraries are solely dependent on government funding but their services and functioning affects due to shrinking budgets, increase in cost of books and other materials, devaluation of rupee conversion of foreign currency rates and many other reasons.

Under the prevailing circumstances, we have to explore ways and means to increase revenue generation so as to reduce dependence and this becomes all the more important in the context of globalization. It is, therefore, essential that avenues are to be explored and adequate thought is given to finding resources for digital libraries and their maintenance.

### ***Prevention of unauthorized utilization***

Common laws and procedures are to be adopted by every government so as to eliminate misuse and theft of information. Filtering systems should be used so that genuine users are not put to inconvenience. With the advent of so-called third industrial revolution, skills and knowledge have become the only source of sustainable long-term competitive advantage. In this global economy, a global system of intellectual property rights is required. Since, the library and information industry are the only channel for disseminating information to all, they have a joint responsibility in fighting against

intellectual property infringement, as it affects every one including users.

The DMCA (Digital Millennium Copyright Act, 1998) also enacted in 1998 offers major advantages for libraries, generally in terms of making digital archival copies.

The protection of IPR (intellectual property rights) and privacy are the two main thrust areas, which WIPO is seriously concerned with. Its December 1996 Diplomatic Conference produced two treaties: The WIPO Copyright Treaty and WIPO Treaty on Performances and Phonogram. The first copyright treaty supplements the Berne convention (Berne Convention) by clarifying issues about digital transmission and distribution of literary or artistic work. Second treaty on performances and phonogram includes a harmonization provision because of the first global effort to protect the exploitation of sound recordings. Talking about protection of intellectual property laws, the US Commissioner of Patents and Trademarks Bruce Lehman commented that it is only a new beginning of Internet and international electronic-commerce development. There is a huge potential for growth and new technologies that have not been used, precisely because of inadequate intellectual property protection globally. One of the main tensions in international trade is related to intellectual property.

The new global digital network environment creates special problems and strains on the traditional system of copyright. One of the main constraints is that national boundaries have become almost meaningless in the digital world. It is difficult to recognize a work of authorship distributed on the internet and it is almost impossible for an individual nation to assert any kind of national control over the product or to protect the IPR that are represented by it.

### ***Technological gap***

Adapting quickly to the new technologies, the electronic revolution combined with improvements in communication makes it imperative to look beyond today and preparing for tomorrow. Changing print material into digital form is a difficult task. In the new millennium, we will close the gaps between the print and digitalized documents, document libraries and knowledge based libraries, IT specialist and information specialist, and information seeker and information providers.

In order to promote continuous upgrading of digital libraries, each country should set up a separate coordinating committee to interact with various agencies, especially computer and telecommunication professional. India, in particular, where technologies are operating together, needs a complete transformation in the information policies to have the way to the digital library era.

### ***Variety of formats***

Digital images may exist in various formats on different computerized networks. Thus, collection, organization, indexing, searching and analysis of such diverse information formats create unique technical challenges (Liu, 2004).

### ***Copyright issue***

Copyright and intellectual property laws are other major issues in digitizing library collections. While undertaking a digitization project, libraries need to decide whether or not the digitized material be protected by copyright law or made available in public domain to be used without taking permission. Other legal issues involved are access and security, privacy and confidentiality, business model and licensing, and so on.



***Technological obsolescence***

Technological advancements are so fast that Digital Libraries need regular up gradation of their software and hardware to be compatible for resources sharing, accessing resources, preserving materials or offering various services.

***Financial issue***

Digital projects are expensive. They require hardware, software and trained staff to take up jobs like scanning, performing quality control, and creating metadata (Lopatin 2006). Besides this, for making the library services available free or at least affordable cost, different pricing models would need to be developed.

***Cultural heritage***

Many digital library and museum collections contain artefacts that are fragile, precious, and of great historical significance. However, the selection and digitization process of such material has not been easy, for technical, organizational and economic reasons (Chen 2004).

***Linguistics and cultural diversity***

India is a multilingual nation with places where more than one language is spoken. The required information might be available in other languages (Indic or foreign). Thus, we need fast and effective translation mechanisms to get over this barrier (Ghosh 2005).

***Archiving issue***

Ensuring the long-term accessibility to digital documents is a global issue. It is still a question as to how digital publication, works of art, image and sound documents and primary data must be archived so that they remain permanently readable and thus, accessible in future (Gastinger 2006).

***Outsourcing***

Lack of sufficient skilled and trained staff to work in changing technological environment

is forcing most libraries for outsourcing digitization work, to be handled by contractual agencies, which challenges information organization and services.

***Digital divide***

Digital libraries are accessible only by those who have availability of necessary infrastructure, including required hardware and software. This makes the information rich person more richer leaving behind an information-poor community.

***Digital reading***

Digital information can't be read as easily as we read printed information according to our convenience. Suitable infrastructure is needed to access the digital resources. Moreover, reading a long article directly from screen poses different problems.

***Lack of cooperation***

There is a lack of strong and sustainable working relationships among libraries and information centres, entering into such venture.

***Lack of national policy***

So far, no policy has been framed keeping in view the legal restrictions to digitized resources. The Government of India is working on it but it may take time to form and put this policy into action, which require huge investments and efforts.

***Political issues***

We are living in a highly political environment where democracy is a 'numbers game' in which all political parties try to tailor their social commitments to suit their own interest groups. In many cases, high profile decisions aim to woo a particular part of the electorate - they are taken with vested interests in mind and may not be for the benefit of the people for whom the underlying policy has been designed.

## Significant library resources digitization initiatives in India

### *Initiatives at the government level*

The union government and state governments have taken considerable initiatives towards the development of digital libraries in the country. The central government commitment towards digital library policy issues is envisaged in the Long Term National IT Policy (National Task Force on Information Technology and National Development 1999) which reflects government's responsibility of providing 'information' to users' in digital form.

Accordingly, the responsibility of envisioning, developing and sustaining hybrid and virtual library and information systems and services rests on library and information progression.

### *Digital libraries of India*

There is a mission to create a portal for the DLI (digital libraries of India) piloted by the Office of the Principal Scientific Advisor to the Government of India, MCIT (Ministry of Communication and Information Technology) with IISc, Bangalore, and Carnegie Mellon University, USA, as partners.

The mission is to create a portal for the digital library initiatives which will foster creativity and free access to all human knowledge. As a first step in realizing this mission, it is proposed to create the digital libraries with a free-to-read, searchable collection of one million books, predominantly in Indian languages, available to everyone over the Internet. This portal will also become an aggregator of all the knowledge and digital contents created by other digital libraries initiatives in India. This portal would provide a gateway to Indian digital libraries in science, arts, culture, music, movies, traditional medicine, palm leaves and many more. The result will be a unique resource accessible to anyone in the world 24×7, without regard to socio-economic

background or nationality.

One of the goals of the Digital Library Initiatives is to provide support for full text indexing and searching facilities based on OCR technologies, where available. The availability of on-line search allows users to locate relevant information quickly and reliably thus, enhancing students' success in their research endeavours. This 24×7 resource would also provide an excellent test bed for language processing research in areas such as machine translation, OCR, summarization, speech and handwriting recognition, intelligent indexing, and information retrieval in Indian languages.

It is expected that the Digital Library Initiatives will be mirrored at several locations worldwide so as to protect the integrity and availability of the data. The Digital Library Initiatives will also partner with other country specific Digital Library initiatives as part of the Universal Library Project ([www.ulib.org](http://www.ulib.org)).

The current status of scanning centres of Digital Library Initiatives shows that approximately 24.521 million pages have been scanned at various centres in India. It shows that approximately 80,240 books have been scanned (Table 1).

### *Vidyanidhi*

Vidyanidhi began as a pilot project in the year 2000 with support from the NISSAT (National Information System for Science and Technology) and the Government of India's DSIR (Department of Scientific and Industrial Research). Vidyanidhi demonstrated the feasibility of e-theses programmes in India. With support from the Ford Foundation and also from Microsoft India, Vidyanidhi is evolving as a national initiative.

Vidyanidhi, in Sanskrit means 'treasure of knowledge' is India's premier digital library initiative to facilitate the creation, archiving and accessing of doctoral theses. Vidyanidhi is an information infrastructure, a Digital

**Table 1** Number of books and pages scanned by language

Language	Number of books	Number of pages
English	49 674	17 456 701
Telugu	14 063	2 804 200
Hindi	6318	1 458 386
Urdu	2977	719 523
Tamil	1550	396 021
Sanskrit	1485	510 340
Persian	1000	245 727
Multi	699	198 158
Others	704	220 350
Arabic	617	198 552
Kannada	586	140 585
Marathi	363	90 947
German	121	46 803
French	63	24 770
Oriya	5	6978
Bengali	4	713
Spanish	4	17
Italian	3	1172
Greek	1	539
Gujarati	1	52
Irish	1	538
Russian	1	73

- Make theses available online (as per the restrictions desired by the doctoral students) and help enhance the visibility of Indian doctoral research.
- Offer tools and resources to strengthen and augment the research capacities of doctoral students and universities.
- Enhance the quality of doctoral research in India by developing and using standard formats and templates.
- Mould 'best practices' in scholarship and scholarly writing among students.
- Prepare the doctoral students in e-publishing, e-scholarship and digital library by offering training programmes and online tutorials.

The Vidyanidhi Digital Library is conceived as having two layers-metadata database and full text of theses. They are currently focusing on building both these layers. Currently it has nearly 50 000 records in the metadata database and 300 full text theses. The Vidyanidhi metadata database is a truly multilingual database with records in English and Indian languages as well. Vidyanidhi has implemented the Unicode standard for Indian languages and scripts.

**Traditional knowledge digital library**

TKDL (Traditional Knowledge Digital Library) is a collaborative project between the NISCAIR (National Institute of Science Communication and Information Resources), the CSIR (Council of Scientific and Industrial Research) of the Ministry of Science and Technology and the AYUSH (Department of Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy) of the Ministry of Health and Family Welfare, which is being implemented at the NISCAIR, New Delhi. An inter-disciplinary team of traditional medicine (Ayurveda, Unani, Siddha, Yoga) experts, patent examiners, IT experts, scientists and technical officers are involved in creation of

Library, a portal of resources, tools and facilities for doctoral research in India. Vidyanidhi is envisioned to evolve as a national repository and a consortium for e-theses through participation and partnership with universities, academic institutions and other stakeholders. Vidyanidhi enhances access to Indian theses and enlarges the reach and audience for Indian doctoral research works.

The following are mission of Vidyanidhi project.

- Develop a repository for Indian doctoral theses.
- Digitize, archive and improve access to doctoral theses in India.

TKDL for Indian systems of medicine.

The project TKDL involves documentation of the knowledge available in public domain on traditional knowledge from the existing literature related to Ayurveda, Unani and Siddha, in digitized format in five international languages which are English, German, French, Japanese and Spanish. TKRC (Traditional Knowledge Resource Classification), an innovative structured classification system for the purpose of systematic arrangement, dissemination and retrieval has been evolved for about 10,500 subgroups against one group in IPC (International Patent Classification), for example, AK61K35/78 related to medicinal plants. Present status of TKDL is reflected in Table 2.

**Table 2** Present status of traditional knowledge digital library

Discipline	Target (number of formulations)	Achieved
Ayurveda	59 000	59 000
Unani	77 000	51 000
Siddha	10 000	—
Yoga	1 500	—
Total	147 500	110 000

At present, TKDL contains 11.0 million pages of information in five international languages.

### Gyandoot

Gyandoot (meaning ‘messenger of knowledge’) is a new intranet-based digital library in the Dhar district, Madhya Pradesh, India, connecting rural public cybercafes. A corresponding website is an extension of Gyandoot intranet providing global access via a portal. Gyandoot was conceived in a discussion with the Secretary, Information Technology, Government of Madhya Pradesh

on 11 November 1999. The pilot project was launched on 29 November 1999 and it was officially commissioned on January 1, 2000. Thus, from concept to commissioning, the entire Gyandoot Digital Library project was executed in the short space of 51 days for a cost of \$57 000.

Gyandoot is a unique form of government to citizen (G2C) digital library activity to address the hardship imposed by transaction costs associated with government services. Located in central India, agriculture and industry are the twin mainstays of business. Close to four thousand million rupees (\$90 million) worth of agricultural commodities are transacted annually, principally soya, cotton, and wheat. Indore is the largest automobile centre in Asia. The local elected governing council (District Panchayat, Dhar) is enabling over half a million rural citizens, the affordable access to various government and market-related needs through state-of-the-art IT kiosks.

### Samadhan kendras

In order to turn food producers/consumers into information producers/consumers, the Indian government is making efforts for the establishment of ‘Samadhan Kendras’ (SKF Rural Support Centres) and ‘Soochana Gumtis’ (SG-Information Kiosks) in the list of industries eligible for loans under various programmes. Digital libraries are being used for the public grievances redressal systems of the state governments through SG facilitation counters in government offices.

Samadhan Kendras/Soochanalaya Gumtis: In Madhya Pradesh, a regional network connects 21 rural cybercafes called Soochanalayas. Each Soochanalaya provides services for about 10–15 village (Gram) Panchayat (a cluster of 20–30 villages) with an aggregate population of 20 000–30 000 people. The network covers 5 of 13 blocks and 3 of 7 tehsils in the district. Soochanalayas are

located at block headquarters, haat (main bazaars, villages and bus depot centres. The Soochanalayas are located on the roadside of the central villages where people normally travel. They together serve a population of over half a million.

A local (matriculation pass) operator called a 'soochak' operates the cybercafes. The soochak is not a government employee but rather a local volunteer. The soochak takes out a bank loan to buy the computer, modem and printer while the Panchayat covers the cost of the phone line. Villagers pay a nominal amount for each service, usually less than 50 cents for any service and there is a set price list. The soochak keeps 90% of the fees with a 10% commission passed back to the Panchayat for new service development and increasing system capacity. With the soochak approach, cybercafes are providing self-employment through entrepreneurship to local rural youth.

#### Indian National Digital Library in Engineering Sciences and Technology (INDEST-AICTE) Consortium

The MHRD (Ministry of Human Resource Development) has setup the INDEST-AICTE consortium on the recommendation made by the expert group appointed by the ministry under the chairmanship of the Prof. N Balakrishnan. The Ministry provides required funds for subscription to electronic resources for 38 institutions including IISc, IITs, NITs, IIMs and a few other centrally funded government institutions through the consortium headquarters set-up at the IIT Delhi. Also, 44 government or government-aided engineering colleges and technical departments in universities have joined the consortium with financial support from the AICTE (All India Council for Technical Education). Moreover, the INDEST-AICTE Consortium, as an open-ended proposition, welcomes other institutions to join it on their

own for sharing benefits it offers in terms of highly discounted rates of subscription and better terms of agreement with the publishers. All electronic resources being subscribed are available from the publishers' websites. The consortium has an active mailing list and a website hosted at the IIT, Delhi.

The INDEST-AICTE consortium is the most ambitious initiative taken so far in the country. The benefit of consortium-based subscriptions to electronic resources is not confined to 38 major technological institutions in the country but is also extended to all AICTE-accredited and UGC-affiliated institutions.

INDEST-AICTE consortium presently subscribes full text databases and e-journals from (ACM) Association for Computing Machinery DL, ASCE (American Society for Civil Engineers) journals, ASME (American Society of Mechanical Engineers) journals, AMR (Academy of Management Review), Capitaline, Euromonitor, GMID (Global Market Information Database), IEL (Institute for Educational Leadership) Online, Indian Standards, Nature, ProQuest Science, Science Direct, Springer Link and bibliographic databases Compendex, Inspec, and MathSciNet.

#### CSIR Consortium

NISCAIR is the nodal organization for developing a consortium for the CSIR laboratories for accessing e-journals. The activities shall range from creation to monitoring of the access facilities of scientific periodicals published by leading international institutions. To start with, an agreement has been signed with Elsevier Science for a period of four years for 1200 journals. Under this scheme, CSIR scientists shall be able to access these journals and download material for their use. Such access to worldwide journal resources playing a very vital role and strengthening research and development in

CSIR laboratories, thus leading to knowledge generation useful for socio-economic development of the country. The objectives are as follows.

- To strengthen the pooling, sharing and electronic access to the CSIR library resources.
- To provide access to world S&T literature to CSIR labs.
- To nucleate the culture of electronic access resulting into evolution of digital libraries.

To date, CSIR has entered into agreement with 11 publishers to access about 3316 international journals across the labs.

#### ***Initiatives at national level institutions***

##### **Parliament library**

A Digital Library has been set up in the computer centre to cater to the needs of members of Parliament, and officers and staff of Lok Sabha Secretariat. A large number of index-based databases of information generated within the Parliament which cater to the instant reference needs of members, officers and research and reference personnel were initially developed by the computer centre. The data stored and available now in PARLIS databases for online retrieval relates to the following.

- Parliamentary questions (full texts of questions and answers since February 2000; indexes from 1985 to 2000 are also available).
- Parliamentary proceedings other than questions (full text of floor versions since the winter session of 1993; indexes from 1985 to 1993 are also available).
- Government and private members' bills from 1985 onwards (only indexes).
- Directions, decisions and observations from the chair, from 1952 onwards.
- President's rule in the states and union territories from 1951.

- Members of council of ministers from 1947 onwards.
- Obituary references made in the houses since provisional parliament.
- Library management functions such as acquisition, processing, issue and return of books have also been computerized using the software package named 'LIBSYS'. A Web-based library catalogue can also be accessed through Internet.
- Documentation service (from 1989 onwards) (important articles published in books, reports, periodicals and newspapers are indexed and annotated and can be accessed through Internet).

#### ***Initiatives at academic institutions of national importance***

##### **IIT, DELHI**

The commitment to Digital Library initiatives and the emphasis upon Web-based digitized collections at the Central Library, IIT Delhi commenced in 1998 with the installation of a fibre optics based campus LAN connected to a 2 Mbps VSNL radio link enabling faster Internet access for the academic community of the Institute. The availability of the high-speed Internet connection has led to the launching of a number of sponsored and un-sponsored projects for developing network-based digitized collections at the Central Library, IIT, Delhi.

IITs are fortunate enough to receive generous grants and projects from government bodies such as AICTE (All India Council for Technical Education) and the MHRD (Ministry of Human Resources Development) to develop their Digital Libraries. A number of online coursewares have been developed. Digitization of old volumes of journals at IIT Delhi is just one example of projects supported by the government.



**Indian Institute of Science, Bangalore**

A project proposal for NSF support under the Indo-US Science and Technology Collaboration initiative has been made by IISc. The IISc, Bangalore would act as a nodal agency to coordinate amongst various academic institutions and governmental agencies from the Indian side. The Carnegie Mellon University would play the same role from the US side.

The aim of the project is to digitize around a million books in the next three years. This joint initiative is planned to synergistically capitalize on the availability of the state-of-the-art of hardware and software in the US for digitizing, storing and accessing of information and the quality personnel available in India. This would act as a forerunner for many such initiatives with other countries, particularly in China and Korea, and would culminate in the grandiose vision of digitizing all the formal knowledge and make available in a location and time independent way for the benefit of the mankind. In order to take a million books to the Web, it is estimated that around 1000 man-years would be needed. If the project is carried out in a developed country like the United States of America, it would cost, at least, around \$40 million besides the cost of the hardware, space and energy.

**National Institute of Technology, Calicut**

Nalanda, the Digital Library initiated in 1999 at the National Institute of Technology, Calicut, is one of the largest Digital Libraries in the country. Nalanda serves members of the campus in meeting their academic and research needs by providing timely and up-to-date information with value added services in all the areas of science, engineering, and technology. Apart from the Digital Libraries reading room, members can access the Nalanda from the entire campus.

**NISCAIR (formerly INSDOC)**

NISCAIR is slowly shifting to electronic libraries that will eventually lead to the establishment of Digital Libraries. With decreasing shelf space and ever growing collections in the libraries, NISCAIR has been advocating the conversion of automated libraries into electronic libraries.

NISCAIR has access to international databases. Information is obtained through online searching from over 1500 international databases. Skilled personnel at NISCAIR perform searches for research scientists and the corporate sector who use these databases for the latest R&D, commercial and market information. National Science Library of NISCAIR has an Electronic Library Division with a rich collection of more than 5000 foreign journals, conference proceedings, and so on, and a large number of databases on CD-ROMs. NISCAIR is the nodal agency for developing a consortium for CSIR laboratories for accessing e-journals. The activity shall range from creation to monitoring of the access facility of scientific periodicals published by leading international institutions.

**National Tuberculosis Institute, Bangalore**

On 28 October 2003, the National Tuberculosis Institute, Bangalore under the initiative and with the support of the Health Inter Network Project, India - TB, launched a digital library. This digital library is comprised of CDs on TB (tuberculosis), available as ready reference tools for programme workers at the District and Primary Health Centre Levels. The CDs on TB have relevant RNTCP (Revised National Tuberculosis Control Programme) documents and scientific literature on programme, treatment, drug resistance and control aspects of TB.

### ***Digitization of art and culture***

#### **Centre for Development of Advanced Computing (C-DAC)-Digital Library of Art Masterpieces**

This is the first initiative of its kind in Asia and it will digitize 200 rare paintings of Rabindranath Tagore and Amrita Shergill from NGMA (National Gallery of Modern Arts). A digital library will be created using the tool DLAS (Digital Library Application Suite), developed by the Digital Library Group, to make the art accessible to a global audience via the World Wide Web. The infrastructure to host this Digital Library would be located at the C-DAC Bangalore.

C-DAC and Hewlett Packard launched the joint initiative 'When Art Meets Technology' for digital preservation, restoration and dissemination of art from the NGMA at Bangalore on 4 February 2003.

#### **Indira Gandhi National Centre for the Arts (IGNCA)-Kalasampada**

IGNCA has taken up the Kalasampada Digital Library-Resource for Indian Cultural Heritage (DL-RICH) project which is sponsored by MCIT.

This project aims to use multimedia computer technology to develop a software package that integrates a variety of cultural information and helps the users to interact and explore the subjects available in image, audio, text, graphics, animation and video on a computer in a non-linear mode, by a click of mouse.

Kalasampada, a unique project of its kind, will facilitate the students, scholars, artists and the research and scientific community to access and view the materials. These materials include several hundred thousand manuscripts, over a hundred thousand slides, thousands of rare books, photographs, audio and video along with highly researched

publications of IGNCA, all accessible from a single window.

The system aims to be a digital repository of content and information with a user-friendly interface. The knowledge base created will help the scholars to explore and visualize the information stored in multiple layers. This will provide a new dimension in the study of the Indian art and culture, in an integrated way, while giving due importance to each medium.

### ***Initiatives within society level organizations***

#### **Mobile DL (Dware dware gyan sampada)**

This is a product from C-DAC Noida. The mission of the project is an Internet-enabled mobile Digital Library brought to common citizens with the purpose of spreading literacy. C-DAC Noida (Department of IT, MCIT) contributes to bringing digitized books to the doorsteps of common citizens. It makes use of a mobile van with a satellite connection for connectivity to the Internet. The van is fitted with a printer, scorer, cutter and binding machine for providing bound books to the end user. Different places, such as schools in villages and other remote areas, will be covered under this programme to promote literacy and demonstrate the use of technology for the masses. The schedule of visits of the mobile Digital Library is made available on their website. Books formatted for book printing may be selected from the website by language, author and title. There are about 350 books in Hindi and English which will be available for download through this website. The site is bilingual (English and Hindi).

### **Conclusion**

Advent of Information Communication Technology has opened new floodgates for

preservation and conservation of library collection. The new paradigm such as digital library not only guaranteed the longevity of resources for future generations but also helps the library and information science professionals to keep pace with technological advancement and provide state of the art services to the users. India is strongly marching forward to remain in the race and many initiatives have been taken by the government and private sectors in developing the digital libraries for present survival and future prosperity.

## References

- American Library Association  
Details available at <<http://www.ala.org/>>
- Ashman A. 2003  
***Digitization: International Encyclopedia of Information Science***, 2nd ed  
edited by John Feather and Paul Sturges  
London: Routledge Taylor and Francis Group, p. 138
- Berne Convention for the Protection of literary and artistic works (Paris Text 19071).  
Details available at <<http://www.law.cornell.edu/treaties/berne/overview.html>>
- Borgman C. 2000  
***From Gutenberg to the global information infrastructure: access to information in the networked world***  
New York: ACM Press
- Chen H. 2006  
**Digital Library Research in the US: an overview with a knowledge management perspective programme**  
*Electronic Library and Information Systems*, 38(3)  
Details available at <[www.emeraldinsight.com](http://www.emeraldinsight.com)> last accessed on 28 September 2006
- Deegan M and Tanner S. 2002  
***Digital Futures: strategies for the information age***  
London: Library Association Publishing.
- Gastinger A. 2006  
**A Report on the Eighth International Bielefeld Conference**  
Academic Library and Information Services – New Paradigms for the Digital Age  
*Library Hi-tech News*. 23(4)  
Details available at <[www.emeraldinsight.com](http://www.emeraldinsight.com)>last accessed on 25 September 2006
- Ghosh, Maitrayee. 2006  
The Public Library System in India: Challenges and Opportunities  
*Library Review* 54(3)  
Details available at <[www.emeraldinsight.com](http://www.emeraldinsight.com)>, last accessed on 29 September 2006
- Gladney H M, Fox E A, Ahmed Z, Ashany R, *et al.* 1994  
**Digital library: gross structure and requirement**  
Details available at <<http://www.csd.tamu.edu/DL94/paper/fox.html>> last accessed on 16 November 2003
- Liu Yan Quan. 2004  
Best Practices Standards and Techniques for Digitizing Libraries: A Snapshot of Library Digitization Practices in the USA.  
*Online Information Review*, 28(5)  
Details available at <[www.emeraldinsight.com](http://www.emeraldinsight.com)>, last accessed on 29 September 2006

Lopatin Laurie. 2006

Library Digitization Projects, Issues and Guidelines: A Survey of Literature Library  
*Hi Tech*. 24(2)

Details available at <[www.emeraldinsight.com](http://www.emeraldinsight.com)>, last accessed on 25 September 2006

Ludwig, Wittegenon

Details available at <<http://www.handprint.com/SC/WIT/wittframe.html>>

Marchionini G and Fox E A. 1999

**Editorial—Progress toward digital libraries: augmentation through integration**

*Information Processing and Management*, 35(3): 219–25

National Task Force on Information Technology and National Development, Government of India. (1999)

**IT action plan (Part III): long term national IT policy. Chapter 5: Content creation industry**

Details available at <<http://it-taskforce.nic.in/actplan3/chap5.htm>>, last accessed on 25 September 2006

Oppenheim C and Smithson D. 1999

**What is the hybrid library?**

*Journal of Information Science*, 25(2): 97–112

Planning Commission.

Ninth Five Year Plan 1997–2002, Vol 1: Development goals, strategy, and policies

New Delhi: Planning Commission, Government of India, p. 256

Pradhan M R. 2004

**Developing Digital Libraries: technologies and challenges**

*Library Herald* 42(2), p. 101

The Digital Millennium Copyright Act. 1998

Details available at <<http://www.loc.gov/copyright/legislation/dmca.pdf>>

United States

UNESCO, IFLA, ICA. 2002

Guidelines for digitization projects for collections and holdings in the public domain, particularly those held by libraries and archives

Details available at <[ifla.org/VII/s19/pubs/digit-guide.pdf](http://ifla.org/VII/s19/pubs/digit-guide.pdf)>, last accessed on 2 September 2004

Witten I H and Bainbridge D. 2003

**How to Build a Digital Library**

Morgan Kaufmann