

---

# UNIT 5 INTRODUCTION TO DIGITAL LIBRARY

---

## Structure

- 5.0 Objectives
- 5.1 Introduction
- 5.2 Concept
- 5.3 Types of Digital Libraries
- 5.4 Major Digital Library Initiatives
- 5.5 Future Trends
- 5.6 Summary
- 5.7 Answers to Self Check Exercises
- 5.8 Keywords
- 5.9 References and Further Reading

---

## 5.0 OBJECTIVES

---

After going through this Unit, you will be able to:

- understand the basic concept, and need for digital libraries;
- explain different types of digitisation; and
- discuss future trends of digital libraries.

---

## 5.1 INTRODUCTION

---

Digital age has brought a tremendous change in the way information is stored and accessed. It is marked by three distinct features: abundance, currency and easy access of information. This has brought about a change in the concept of libraries, their collection and services. Many new terms viz., ‘digital libraries’, libraries without walls’, ‘virtual libraries’ are emerging to describe the libraries of present day age.

The term ‘digital library’ is a shift from the earlier term electronic library which was used for the last two decades to describe the book-less library which relies on telecommunication and computers to provide users with whatever information they need. A digital library is popularly viewed as an electronic version of a library where storage is in digital form, allowing direct communication to obtain material and copying it from a master version. It combines technology and information resources to allow remote access, breaking down the physical barrier between resources. In Wilensky’s view “the digital library will be a collection of distributed information services, producers will make it available, and consumers will find it through the automated agents”. In this model it appears that the traditional libraries will have no role to play. How far this will be true only time can tell.

In the early stages of development of digital libraries the main focus was on providing dial up access to Online Public Access Catalogues (OPAC). The term however evokes different meaning for different people. To some it may simply mean computerisation of the traditional library system. To those with library science

background it means doing things in a new way, using new type of information resources, new approach to acquisition, new methods of storage and preservation, new approaches to classification and cataloguing, new ways of interaction with the patrons with more reliance on electronic system and networks. As it stands today, most libraries in the developed countries have their own homepages providing links to local information, electronic databases, bibliographic as well as full text, apart from its own online system of collection and services.

Digital libraries in future will not be a standalone version. The explosive growth in networked connectivity and rapid advances in computing power are replacing the older notions of standalone information utilities with newer notions of integrated digital libraries. The integrated digital library creates a shared environment linking everything from personal collection, collection of conventional libraries and large databases spread all over the world.

In the recent years the term ‘virtual library’ is becoming more popular. It is being used to describe libraries that provide access to digital information using variety of networks, specifically the internet and the World Wide Web, irrespective of place and time. According to Gilbert “it is an aggregate of libraries or literature bases, the catalogue or bibliographies of which are accessible electronically (e.g. with a personal computer) and of which some may offer document ordering and delivery services. The center of the virtual library is by definition the individual user, or his/her work station”. Thus in the present day context virtual library is the convergence of a number of concepts: electronic browsers, online catalogues and literature bases, and empowerment of the end users.

In Toren and Czech’s view, libraries in future will become icons on the screen and library buildings will function as book warehouses. The future implication of such a situation needs to be contemplated seriously.

---

## 5.2 CONCEPT

---

### Defining Digital Libraries

The term “digital library” is the most recent in a long series of names for a concept that has been written about nearly as long as the development of the first computer: a computerised “library” that would supplement, adds functionality, and even replaces traditional libraries.

In comparison to traditional libraries, digital libraries provide efficient and qualitative services by collecting, organizing, storing, disseminating, retrieving and preserving the information. Digital libraries support preservation besides making information retrieval and delivery more comfortable. It provides online access to historical and cultural documents whose existence is endangered due to physical decay. The major areas which offer digital libraries great exploitation are: Information retrieval, multimedia database, data mining, data warehouse, on-line information repositories, image processing, hypertext, World Wide Web and Wide Area Information Services (WAIS).

Digital libraries necessarily include a strong focus on the management of digital content, just as traditional libraries have focused for long on the management of content in physical forms. Most of the digital content that is being managed includes human language, either in the form of character-coded electronic text, scanned versions of

printed or handwritten text, or digital representations of human speech. Language technology therefore plays a major role in managing digital content. This comes as no surprise, of course. Digital libraries today make good use of what we know about searching large collections, and techniques such as machine-assisted indexing are employed increasingly often as we strive to extend our reach to progressively larger collections. But we are on the verge of a new era, one in which our machines will learn from what we do and then apply those capabilities to enable the management of digital content at a far larger scale than we could ever hope to do ourselves.

Few advantages of digital libraries according to Haddouti are:

- User can access the information anywhere
- Reduces bureaucracy by providing access to the information
- The information is not necessarily located in same place
- Understanding the catalogue structure is not necessary
- Cross references to other documents speed up the work of users
- Full text search
- Protected information source
- Wide exploration and exploitation of the information

The knowledge dissemination is an integral part of success story of popularity of creating digital libraries. The aim is to provide universal access to human knowledge, and given the advancement of digital storage and communications this goal is now achievable.

### **Distributed Models**

Libraries are increasingly adopting distributed models for information access and management, and more often use open and collaborative models for developing library content and services. With the incorporation of open models and distributed technologies, the libraries have the potential to get more involved in knowledge creation, dissemination, and use. In reference to libraries, the creation and dissemination of knowledge—in ways that represent the library's contributions more broadly and that intertwine the library with the other stakeholders in these activities. The library becomes a collaborator within the academy, yet retains its distinct identity.

### **Open Paradigms and Models**

There is new trend emerging as Open Source movement—the concept of collaborative software development with developers sharing the source code — reflects a fundamental shift away from proprietary software and systems. These open models are appearing in new applications areas such as the Open Knowledge Initiative to share learning technologies. The increasing interest in open models is leading towards more generalized acceptance of collaborative development and sharing of intellectual goods and services. Cyber law experts suggest that the creation of a “commons,” wherein the free exchange of ideas and collaboration prevail, is fundamental to an open society. Themes of openness and collaborative exchange have also emerged in the context of publishing, particularly with respect to the relationship between authors and commercial publishers. As information becomes more distributed and open models of exchange become more common, the library's relationship with content creators, publishers, and consumers will change. In these open trends there is evidence of a shift from publication as product to publication as process. When content is available

in such a shape that can be enhanced or supplemented over time, it becomes more dynamic and the “versions” become more cumulative. Few people forecast this shift as the ultimate challenge to current copyright law. Such a shift will have significant impact on organisations whose current role is to manage publications in both traditional and digital forms. As this shift continues, there are likely to be further changes in the library’s information management functions.

In this second phase in the evolution of library roles, the library starts to engage in collaboration as a strategy to address its core mission of building collections, maintaining access, and providing service. As responsibilities for content and services become more distributed, models of central control give way to new mechanisms for coordination and collaboration. Ultimately, the processes of scholarly communication become as critical as traditional publication products.

### Digital Collections Vs Digital Library

In the last decade substantial progress has been made in creating large-scale digital collections. It is extremely important to distinguish digital collections from digital libraries. There is no clear definition about what exactly constitutes a digital library. Digital collections are “raw content,” while “digital libraries [are] the systems that make digital collections come alive, make it usefully accessible, useful for accomplishing work, and connect them with communities.” The collections gain value only when these are surrounded by a matrix of content and interpretation that makes them useful. Therefore it should be ascertained that we develop digital libraries, not just digital collections.

Care should be taken to surround collections with appropriate metadata supplying context and interpretation, to develop synergy. It is the time to “build massive, comprehensive digital collections that scholars, students, and other researchers can use with more ease than they use the book-based collections.”

Three general characteristics of the digital library of the future are:

- A comprehensive collection of resources important for Scholarship, teaching, and learning;
- Readily accessible to all types of users
- Managed and maintained by professionals

The information explosion, the wide bandwidth data networks and the potential of Internet-based technologies - such as the Web - make digital libraries one of the important application areas of computer science.

### Self Check Exercise

**Note:** i) Write your answers in the space given below.

ii) Check your answers with the answers given at the end of this Unit.

1) Discuss three general characteristics of the digital library of the future.

.....  
.....  
.....  
.....  
.....

---

## 5.3 TYPES OF DIGITAL LIBRARIES

---

Digital libraries can be grouped in different ways. They can be classified by origin, such as digital libraries developed in the USA as part of DLI 1 and DLI 2 (the Digital Library Initiatives), digital libraries developed in the course of the eLib (Electronic Libraries) programme in the UK, digital libraries built by individual institutions, digital libraries that are part of national libraries, digital libraries that are part of universities; or by period, by country of origin, and so on.

- early digital libraries, e.g. ELINOR, Gutenberg
- digital libraries of institutional publications, e.g. ACM, IEL
- digital library developments at national libraries, e.g. the British Library, Library of Congress (THOMAS), Digital Library of Canada
- digital libraries at universities, e.g. Berkeley Digital Library SunSITE Bodleian Library Digital Library Projects, California Digital Library, DIGILIB, iGEMS and SETIS
- digital libraries of special materials, e.g. Alexandria, Informedia, Grainger Engineering Library
- digital libraries as research projects, e.g. GDL, NCSTRL, NDLTD
- digital libraries as hybrid library projects, e.g., HeadLine.

### Self Check Exercise

- Note:** i) Write your answers in the space given below.  
ii) Check your answers with the answers given at the end of this Unit.
- 2) Classify different types of digital libraries.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

---

## 5.4 MAJOR DIGITAL LIBRARY INITIATIVES

---

### • The British Library's Digital Libraries Programme

(<http://www.bl.uk/aboutus/stratpolprog/digi/dom/index.html>)

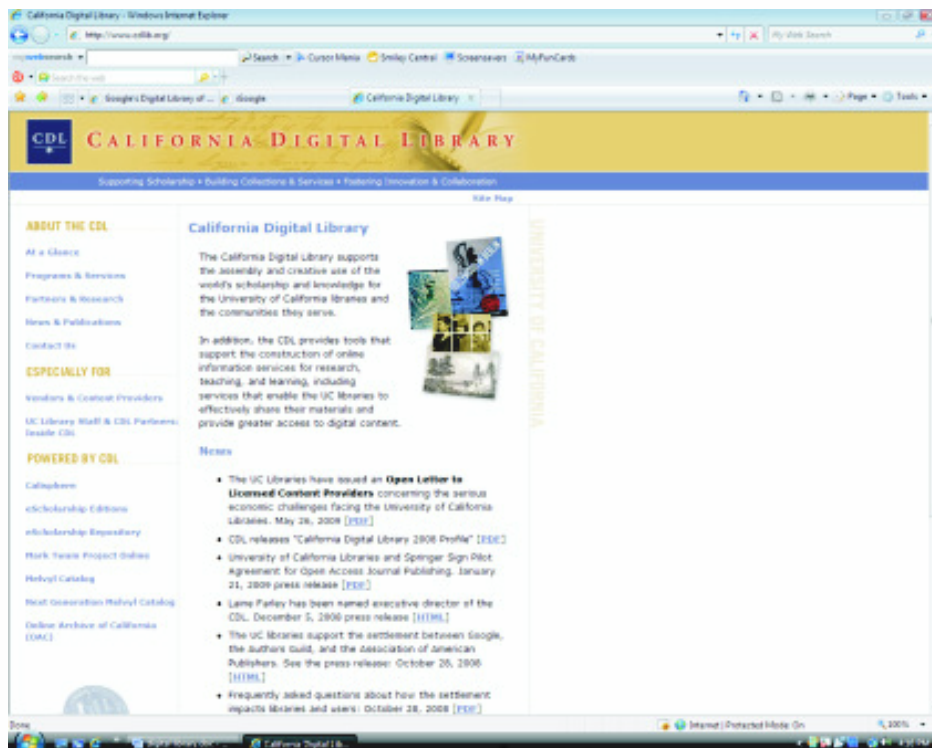
The Digital Libraries Research Programme at British Library Research and Innovation Centre (BLRIC) is establishing a digital library information service based on the British library collections.



- THOMAS - Library of Congress Digital Library** (<http://thomas.loc.gov/>)  
 The Library of Congress Digital Library, Thomas was launched in January 1995, at the inception of the 104<sup>th</sup> Congress to make federal legislative information freely available to the public.



- California Digital Library** (<http://www.cdlib.org/>)  
 The California Digital Library was established in 1997 at the University of California. It supports the University of California libraries in their mission of providing access to the world’s knowledge for the UC campuses and the communities they serve. The CDL also maintains its own distinctive programs emphasizing the development and management of digital collections, innovation in scholarly publishing, and the long-term preservation of digital information.



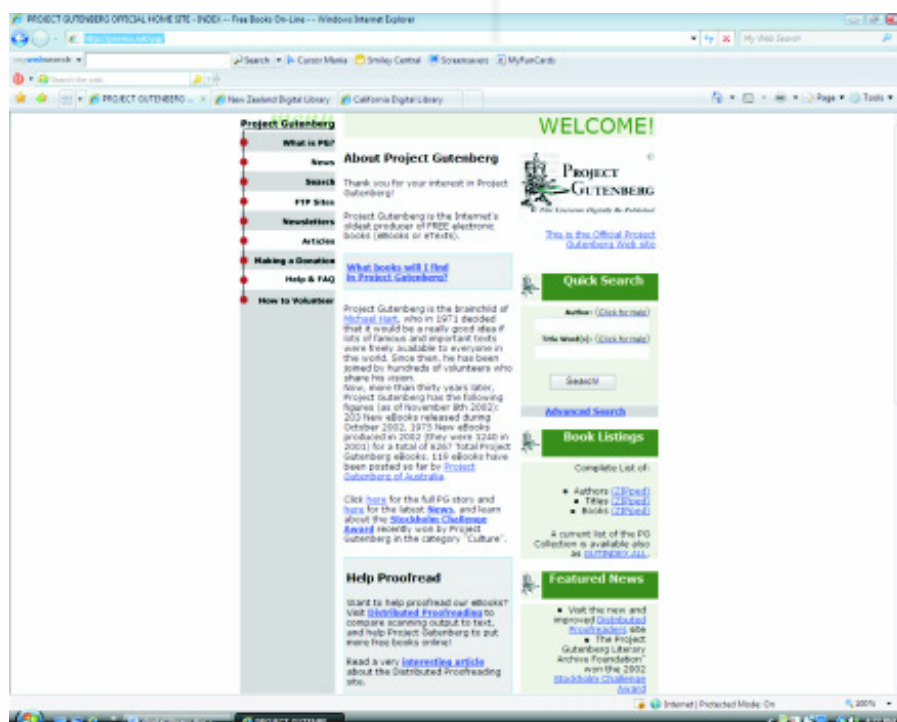
- **Google Digital Library of Alexandria**

Google announced the library scanning project in December 2004. It has four library partners viz. Stanford University, Oxford University, New York Public Library and University of Michigan. The major publishing houses like McGraw-Hill and Penguin Group have sued Google for scanning books without permission.

Reference: <http://googlesystem.blogspot.com/2006/08/googles-digital-library-of-alexandria.html>

- **Gutenberg** (<http://promo.net/pg/>)

The project Gutenberg began in 1971 at the Materials Research Lab, the University of Illinois. The prime objective of this project was to facilitate the world's great literature to electronic versions for the public access.





- **The IEEE Electronic Library**

([http://www.ieee.org/portal/innovate/products/research/ieee\\_iel.html](http://www.ieee.org/portal/innovate/products/research/ieee_iel.html))

The IEEE digital library is the gateway to valuable, cutting-edge research, standards and educational courses with more than two million articles. It offers 100% full-text searchable content with full-page PDF images of all IEEE articles, papers and standards.



- **International Children's Digital Library (ICDL) (<http://en.childrenslibrary.org/>)**

The ICDL was created by an interdisciplinary research team at the University of Maryland in cooperation with the Internet Archives. This was established to create a collection of more than 10,000 books in at least 100 languages that is freely available to children, teachers, librarians, parents, and scholars throughout the world via the Internet.





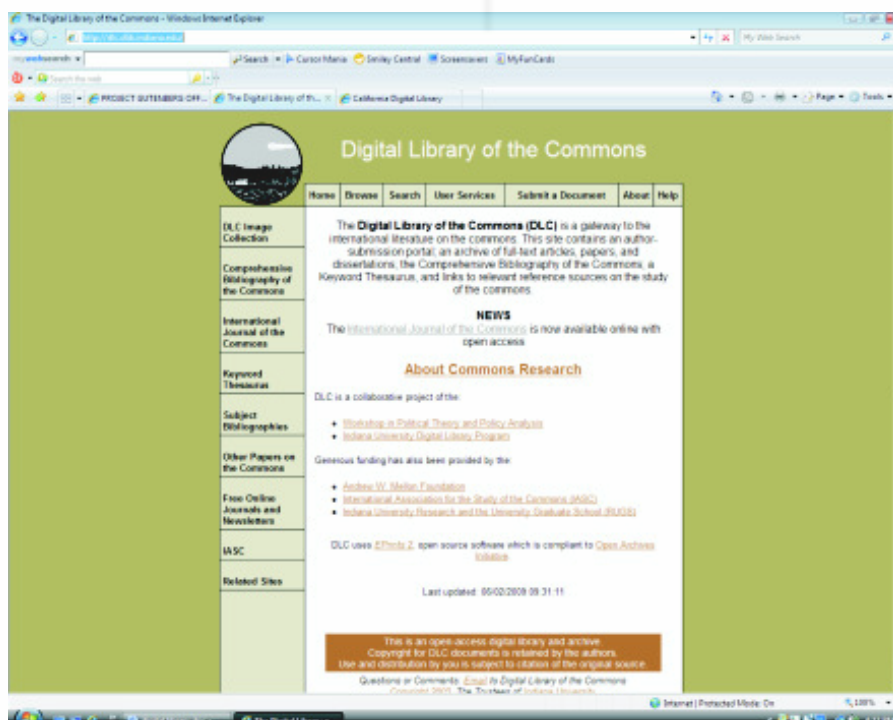
- **The New Zealand Digital Library Project** (<http://nzdl.sadl.uleth.ca/cgi-bin/library.cgi>)

The New Zealand Digital Library Project is a research programme at the University of Waikato. The main objective of this project is to develop the underlying technology for digital libraries and make it available publicly.



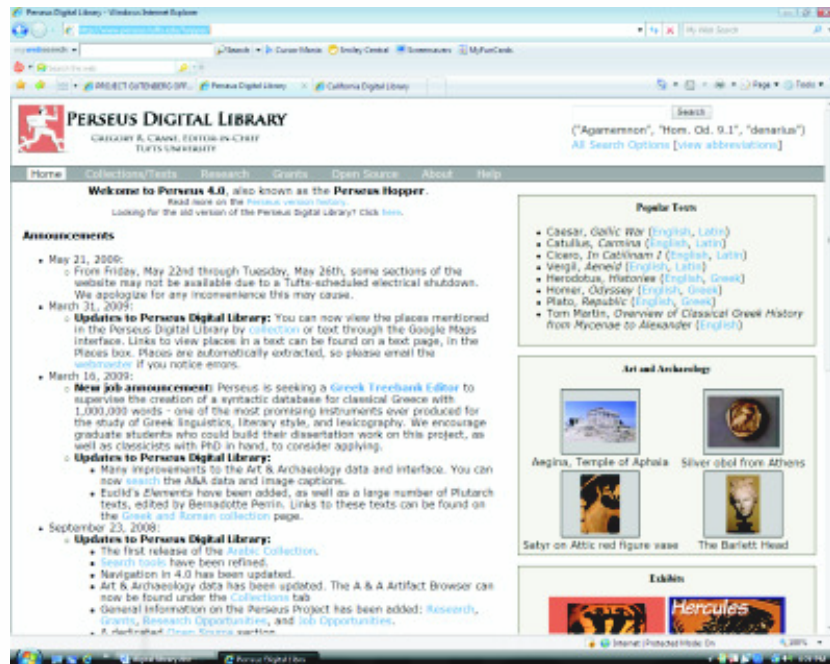
- **Digital Library of the Commons** (<http://dlc.dlib.indiana.edu/>)

The Digital Library of the Commons (DLC) is running on Eprints2, which provides free access to an archive of international literature on the commons, common-pool resources and common property. Features for authors and readers include advanced searching; browsing by region, sector, and author name; an author submission portal for uploading a variety of document formats; and a service that uses email to alert subscribers to new documents in their area of interest.



- **Perseus Digital Library** (<http://www.perseus.tufts.edu/hopper/>)

Perseus is an evolving digital library, to bring a wide range of source materials to as large an audience as possible.



- **The German Digital Library Programme GLOBAL INFO**

The German Digital Library Programme GLOBAL INFO is funded by the federal ministry for education and research from 1998. The main objective of this initiative is to provide optimal access to the world-wide electronic and multimedia information on full texts, literature references, factual databases and software.

Reference: <http://dlib.anu.edu.au/dlib/april99/04rusch-feja.html>

- **The Sydney Electronic Text and Image Service (SETIS)**

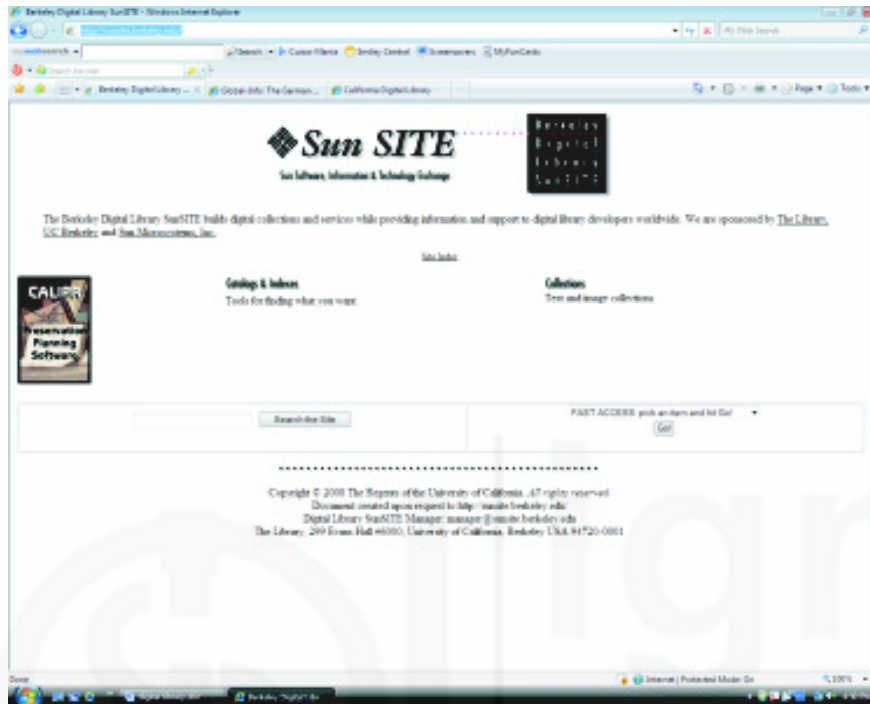
(<http://setis.library.usyd.edu.au/>)

SETIS was launched in 1995 at the University of Sydney. It provides access to a large number of networked and in-house full text databases. It also engaged in a number of text and image creation projects.



- **The Berkeley Digital Library** (<http://sunsite.berkeley.edu/>)

The Berkeley Digital Library project began as an inter-agency, academic teaming to research collaboration techniques. It continues and is currently developing the tools and technologies to support highly improved models of the “scholarly information life cycle”. The goal is to facilitate the move from the current centralized, discrete publishing model, to a distributed continuous, and self-publishing model. It provide access to a large variety of scholarly publications.



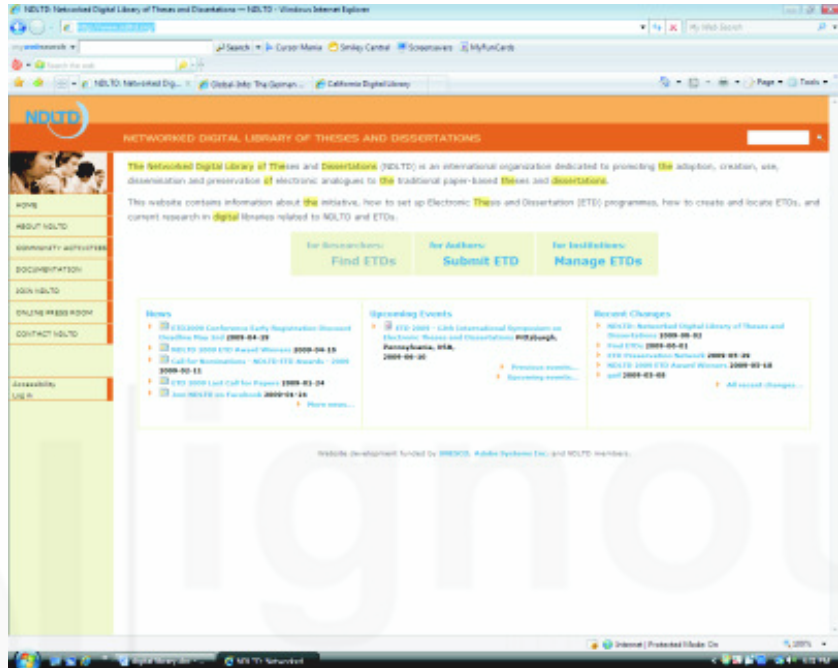
- **Informedia Digital Video Library** (<http://www.informedia.cs.cmu.edu/>)

This is a project at Carnegie Mellon University and the overarching goal of the Informedia initiative is to achieve machine understanding of video and film media, including all aspects of search, retrieval, visualization and summarization in both contemporaneous and archival content collections. The Informedia-II seeks to improve the dynamic extraction, summarization, visualization and presentation of distributed video.



- **The Networked Digital Library of Theses and Dissertations (NDLTD)** (<http://www.ndltd.org/>)

The Networked Digital Library of Theses and Dissertations is an international organisation dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations. This contains information about the initiative, how to set up Electronic Thesis and Dissertation (ETD) programmes, how to create and locate ETDs, and current research in digital libraries related to NDLTD and ETDs.



- **The Bradman Digital Library, Australia** (<http://www.slsa.sa.gov.au/bradman/>)

This digital library was created to give world wide access to collection of memorabilia devoted to Sir Don Bradman and held by the Mortlock State Library of South Australia. It contains biographical information about Bradman, a digital exhibition of artifacts, and a series of scrapbooks covering the years 1925-26 to 1948-49, containing press cuttings, notes and photographs.





- **The University of Adelaide Digital Library** (<http://digital.library.adelaide.edu.au/>)

The Digital Library undertakes projects aimed at enhancing online access to information for their members. This provides access to exam papers available online, Australian digital theses collection and e-books available at Adelaide.



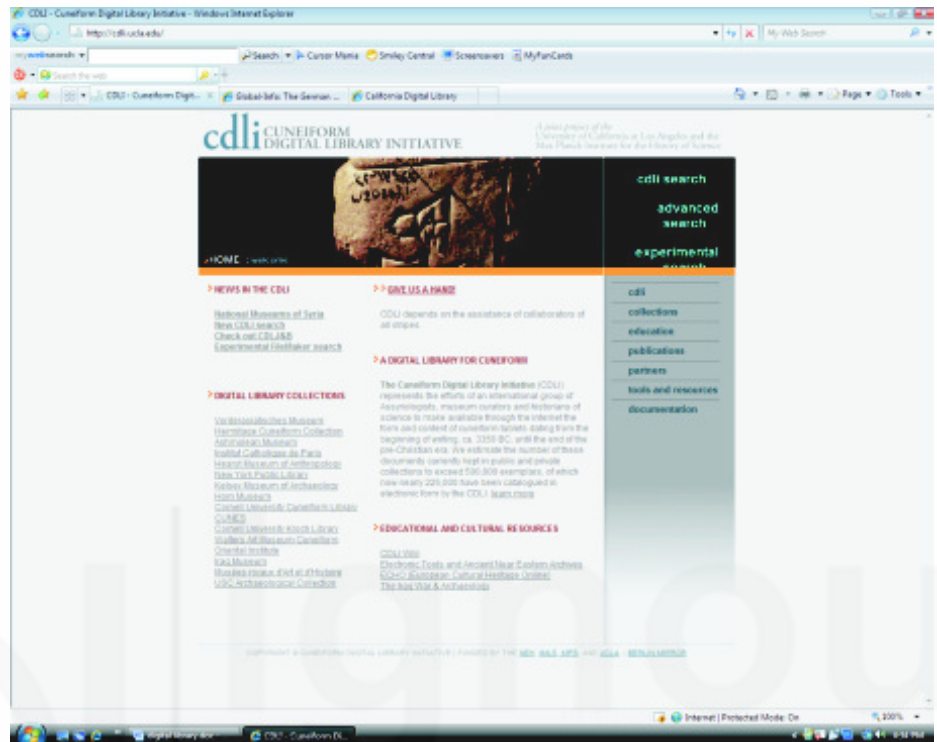
- **National Science Foundation Digital Library** (<http://nsdl.org/>)

The National Science Foundation Digital Library at the University of Texas at Austin is a dynamic archive of information on digital morphology and high-resolution X-ray computed tomography of biological specimens.



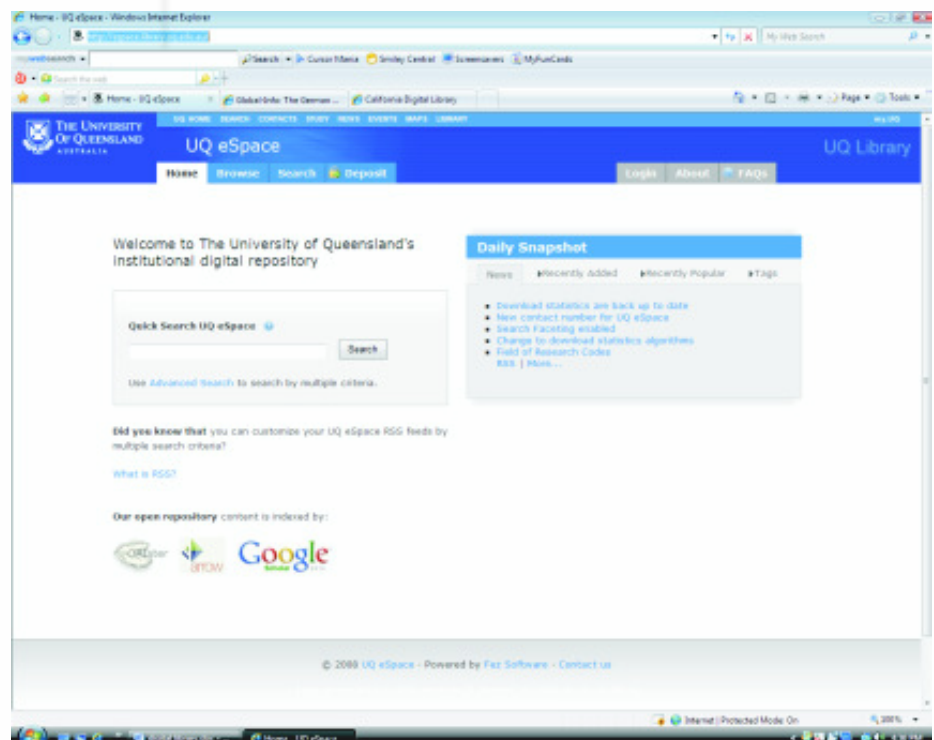
- **The Cuneiform Digital Library Initiative (CDLI) (<http://cdli.ucla.edu/>)**

The Cuneiform Digital Library initiative represents the efforts of an international group of Assyriologists, museum curators and historians of science to make available through the internet the form and content of cuneiform tablets dating from the beginning of writing until the end of the pre-Christian era.



- **UQ eSpace (<http://espace.library.uq.edu.au/>)**

UQ eSpace is the University of Queensland’s institutional digital repository for publications, research, and teaching materials. Deposited material covers a very wide range of subjects and disciplines. This also holds the electronic full text of many peer-reviewed published articles and conference papers, book chapters, theses and other forms of written research from UQ academic staff and students.





- **Traditional Knowledge Digital Library**

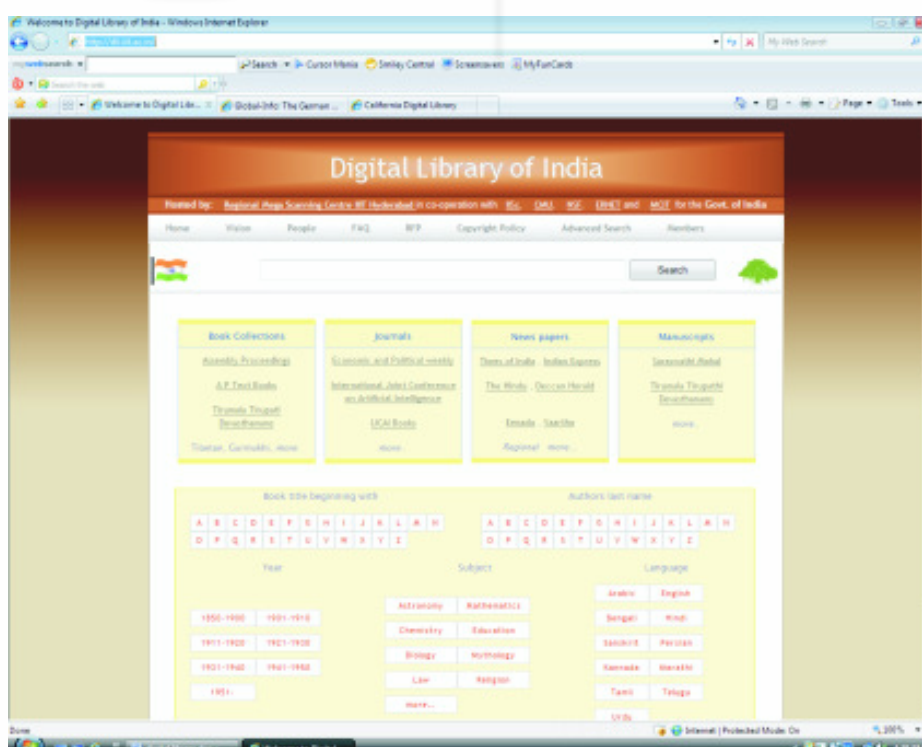
(<http://www.tkdlib.res.in/tkdlib/langdefault/common/home.asp?GL=Eng>)

The Traditional Knowledge Digital Library is a well known Indian digital library initiative being implemented by the National Institute of Science Communication and Information Resources (NISCAIR). The major objective is to provide information on the Indian system of medicine such as Ayurveda, Unani, Siddha, Yoga, Naturopathy and Tribal Medicine.



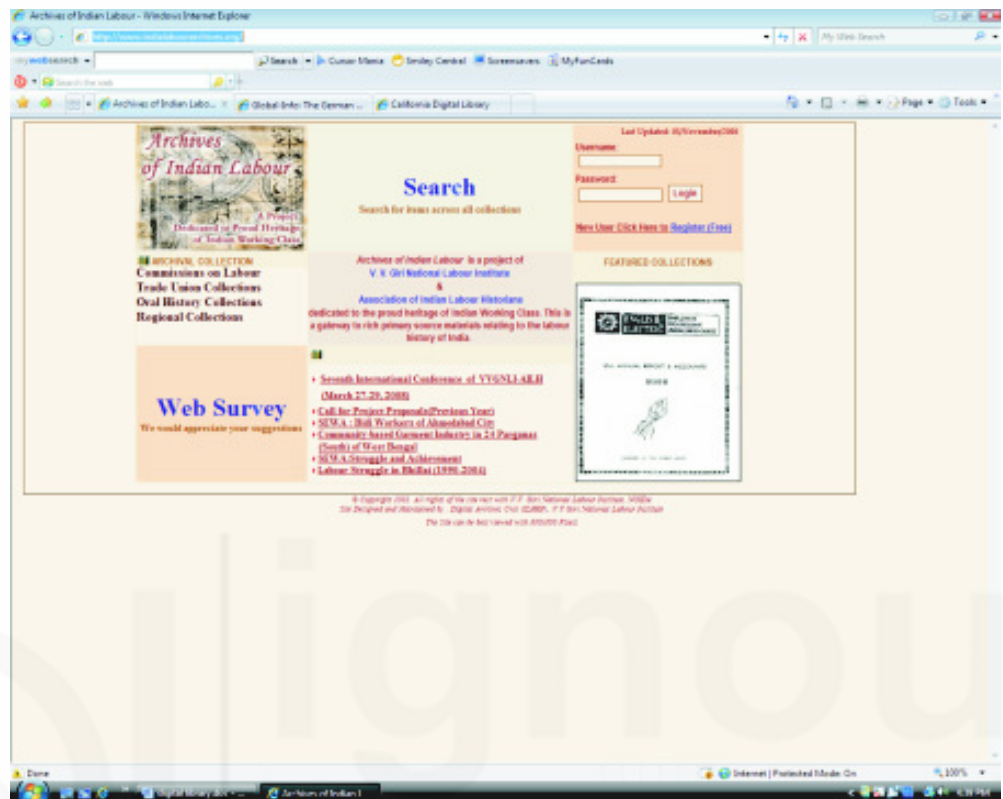
- **The Digital Library of India (DLI)** (<http://dli.iit.ac.in/>)

The Digital Library of India is the greatest digital library initiative in the country. DLI is a part of Universal Digital Library (UDL) and Million Books Projects, coordinated by the Carnegie Mellon University, USA.



- **The Archives of Indian Labour** (<http://www.indialabourarchives.org/>)

The Archives of Indian Labour is a collaborative project of V.V.Giri National Labour Institute and the Association of Indian Labour Historians. The main objective is to preserve and make accessible archival documents on the working class of India.



## 5.5 FUTURE TRENDS

Although the term digital library is used widely in the literature, a new term, ‘hybrid library’, appeared in the course of digital library research in the UK. A hybrid library has been defined as a library where digital and printed information resources co-exist and are brought together in an integrated information service accessible locally as well as remotely (HyLife, 2002a). A number of researcher believe that for the foreseeable future we shall live in the world of hybrid libraries that will integrate traditional libraries with the emerging digital ones (for example, Oppenheim and Smithson, 1999; Pinfield et al., 1998; Rusbridge, 1998). Pinfield at al. (1998) comment that the hybrid library is on the continuum between the conventional and digital library, where electronic and paper-based information sources are used alongside each other. Rusbridge (1998) suggests that a hybrid library brings a range of technologies from different sources together, and integrates systems and services in both the electronic and print environments. He further argues that ‘the name hybrid library is intended to reflect the transitional state of the library, which today can neither be fully print not fully digital’.

There are numerous areas of research related to the historic interests of the digital library community that are at the crossroads of technology and social science and which will demand investment and attention in the coming years; many of these are natural extensions and elaborations of the collaborations initiated by the past decade of digital library research programs. Below mentioned are some of the driving force areas for future of digitisation

- Personal information management. As more and more of the activities in our lives are captured, represented and stored in digital form, the questions of how we organize, manage, share, and preserve these digital representations will become increasingly crucial. Among the trends lending urgency to this research area are the development of digital medical records (in the broadest sense), e-portfolios in the education environment, the overall shift of communications to email, and the amassing of very large personal collections of digital content (text, images, video, sound recordings, etc.)
- Long term relationships between humans and information collections and systems. This is related to personal information management, but also considers evolutionary characteristics of behaviour, systems that learn, personalization, system to system migration across generations of technologies, and similar questions. This is connected to human-computer interface studies and also to studies of how individuals and groups seek, discovers, use and share information, but goes beyond the typical concerns of both to take a very long time horizon perspective.
- Role of digital libraries, digital collections and other information services in supporting teaching, learning, and human development. The analysis here needs to be done not on a relatively transactional basis (i.e. how can a given system support achievement of a specific curricular goal in seventh grade mathematics) but how information resources and services can be partners over development and learning that spans an entire human lifetime, from early childhood to old age.
- Active environments for computer supported collaborative work offer the starting point for another research program. These environments are called for, under the term “colaboratories”, by the various cyber infrastructure and e-science programs, but have much more general applicability for collaboration and social interactions. From one perspective, these environments are natural extensions of digital library environments, but at least some sectors of the digital library community have always found active work environments to be an uncomfortable fit with the rather passive tradition of libraries; perhaps here the baggage of “digital libraries” as the disciplinary frame is less than helpful. But there is a rich research agenda that connects literatures and evidence with authoring, analysis and re-use in a much more comprehensive way than we have done to date; this would consider, for example, the interactions between the practices of scholarly authoring and communication on one hand, and on the other, the shifting practices of scholarship that are being recognized and accelerated by investments in e-science and e-research.

---

## 5.6 SUMMARY

---

Libraries have always played a significant role in society, and digital libraries with the promise of breaking the barriers of geographical distance, language and culture, have a potentially even more significant social role. Digital libraries will not only change our reading and information use habits, they are also going to bring major changes in the economic models of information generation, distribution and management functions.

A tremendous amount of research and development activity has gone into the study of digital libraries. Many issues have been addressed and problems have been partly or fully resolved. Researchers from a variety of disciplines, such as library and

information science, computer science and engineering, social sciences and humanities are working closely together to look into the myriad of unresolved issues.

For exploiting the benefits of Digital Library in Indian languages there is urgent need of tools and applications such as OCRs and Machine Translation systems so that user can take benefit of reading rare classics published in any language and researchers are able to use these tools for their linguistic research. This parallel aligned corpus development is first attempt in context of Indian languages. This is the initiation of several efforts which will follow the trend of enhancing the research in the field of Computational Linguistics. The parallel corpus as a Translation Memory (TM) will be a valuable source in improving the translation system and translators' efficiency.

It will boost the development of Lexical and Terminology databases with the combination of Quantitative and Qualitative Analysis of Text. Text Analyzer is a new kind of tool which is helpful in lexicography, knowledge acquisition, language and writing variation studies. Digital libraries creation have been a good test bed for OCR's and now that the world is moving towards speech to speech translation all these tools together will help building one for Indian languages.

---

## 5.7 ANSWERS TO SELF CHECK EXERCISES

---

- 1) Three general characteristics of the digital library of the future are:
  - A comprehensive collection of resources important for Scholarship, teaching, and learning
  - Readily accessible to all types of users
  - Managed and maintained by professionals.
- 2) Digital libraries can be classified broadly into:
  - early digital libraries, e.g. ELINOR, Gutenberg
  - digital libraries of institutional publications, e.g. ACM, IEL
  - digital library developments at national libraries, e.g. the British Library, Library of Congress (THOMAS), Digital Library of Canada
  - digital libraries at universities, e.g. Berkeley Digital Library SunSITE Bodleian Library Digital Library Projects, California Digital Library, DIGILIB, iGEMS and SETIS
  - digital libraries of special materials, e.g. Alexandria, Informedia, Grainger Engineering Library
  - digital libraries as research projects, e.g. GDL, NCSTRL, NDLTD
  - digital libraries as hybrid library projects, e.g., HeadLine.

---

## 5.8 KEYWORDS

---

- Hybrid library** : Libraries containing a mix of traditional print library resources and the growing number of electronic resources.
- OCR** : Optical Character Recognition, or OCR, is a technology that enables you to convert different types of documents, such as scanned paper documents, PDF files or images captured by a digital camera into editable and searchable data.

**Open Knowledge Initiative** : The Open Knowledge Initiative (O.K.I.) is an open and extensible architecture for learning technology specifically targeted to the needs of the higher education community.

**Open Source Movement** : A broad-reaching movement of individuals who support the use of open source licences for some or all software. Open source software is made available for anybody to use or modify, as its source code is made available.

---

## 5.9 REFERENCES AND FURTHER READING

---

Chowdhury, G G and Chowdhury, Sudatta, (2003) “*Introduction to Digital Libraries*”, Facet Publishing, UK. Print

Haddouti, H. (1997) The Digital Library Initiatives. Proceedings of the Symposium on The Arab World and Information Society Tunis, May 4-8, UNESCO, (invited Talk)

<http://dspace.iimk.ac.in/bitstream/2259/252/1/05-mgs-ps-paper.pdf>

<http://www.dlib.org/dlib/july05/lynch/07lynch.html>

Oppenheim C and Smithson D. (1999) What is the hybrid library? *Journal of Information Science* 25(2):97–112.

Pinfield, Stephen [et al] (1998). “Realizing the Hybrid Library” In *D-Lib Magazine*, October URL:<http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/october98/10pinfield.html>

Rusbridge, Chris (1998) Towards the hybrid library In *D-Lib Magazine*, July-August. <<http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html>>

Wilensky, R. (2000), Digital library resources as a basis for collaborative work. *J. Am. Soc. Inf. Sci.*, 51: 228–245.