New Era Digital Library: Re-engineering Library

Ram Jha **Abhishek Kumar**

This paper focuses on and discusses the evolution, developments, importance and essence of Digital Library and its integration in new era in regard to multimedia data, Internetworking aspects and Hypermedia Applications. Storage, retrieval, cataloguing and design architecture of digitization of library will take the centre stage in the context of fast growing technology and in view of very large users requiring information at lightening speed simultaneously. Also, this paper discusses the reengineering perspective of Library.

Keywords: Hypermedia, Digital Library, OCR

1. History, Evolution & Revolution of Library

The era of B.C. or late 18th century library is considered to be the era of physical collections of data in the form of handwritten manuscripts, typeset print, typewritten, etc. During that period, Libraries used to function like a so called ware house. The information in those libraries consisted of literature, books, manuscripts, papers, publications, etc. from the every walks of life be it humanities, law, sciences (social & political) & serving the needful with tons of repository. But now, with the evolution of Library and Information science, digitization of information, publications, research papers, manuscripts, etc. with the help of growing technology the traditional resources of a library have undergone a massive reformation with the application of universally accepted technology. With the help of these advanced technologies, the physical collection of traditional library is completely and comfortably converted into Digital Library, and further, with the help of multifaceted data and internetworking system, Digital Library are made available to researchers/users universally using Hypermedia technology.

Practically, the digital library era can be said to have begun in early 1990s. As predicted by several authors, it was only then that various technological factors (computer storage, high speed networking, world wide web) came together to enable the digital library to become a reality.

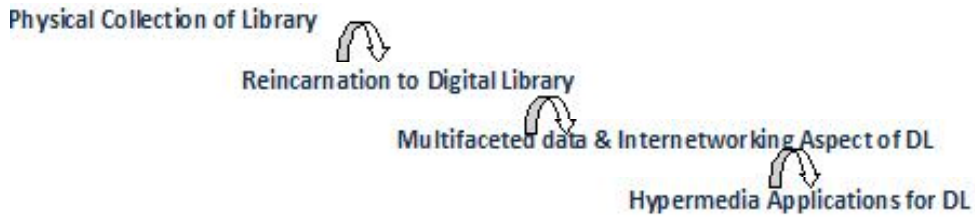
Now & Then of library science reformation

Libraries Then (Early Time)

- Ugrait in Syria excavation revealed library containing diplomatic texts as well poetry, way back to around 1200 BCE.
- 7th century – first systematically assembled library at Nineveh.

- Library of Alexandria best example of an early library in 3rd century BC.
- Cataloguing by library classification & notation in Imperial library (Han Dynasty).
- 19th century - classification system devised on Baconian method which emphasize grouping by subject rather than alphabetical.
- 20th century – the term library science is coined by infusing library & information science amal gamation. In this the philosophy is more tied to day-to-day business of running a library.
- 21st century – Digital library in which collections are stored in digital formats & accessible by computers.
- Hypermedia applications supporting digital library in the middleware services for navigation & annotation which eventually drives the scholarly works.

2. Stages of Transofrmation



The sequential development of Library Science and its perpetual conversion into Library & Information Science are based on the quadrilateral bases which are as follows:

3. Library to Digital Library a Paradigm Shift

Need for Digital Library: With the passage of time from Syria Dynasty to 19th century, there has been a pile of books. Now libraries find it difficult to accommodate these ready material due to scarcity of space. If the problem is confined to walls of only space it would have been managed but the classification of storage, faster retrieval of data started creeping in. The librarians always break rules for devising a new format for catalogue, storage & retrieval to accommodate the current need of the hour.

This is the time when an urgent need is felt for research in the area of library science. In more recent years, with the growth of digital technology & information science the field has significant influence in terms of:-

- Collection and preservation of data.
- Classification of data (cataloguing).
- Organization & Dissemination of information resources.

(Before moving ahead it is important to know about digitization, a procedure to make books in digital format at a high speed than ever before & having a comparatively lower cost for creating a digital library for millions of books.)

4. Digital library

A digital library is a library in which collection are stored in digital format and accessed through computers. The contents may be stored locally or over network and can be accessed remotely via network. A managed environment of multimedia materials in digital form, for the benefit of its user population, structured to facilitate access to its contents, and equipped with aids to navigate the global network…with users and holdings totally distributed, but managed as a coherent whole (Collier, 1997).

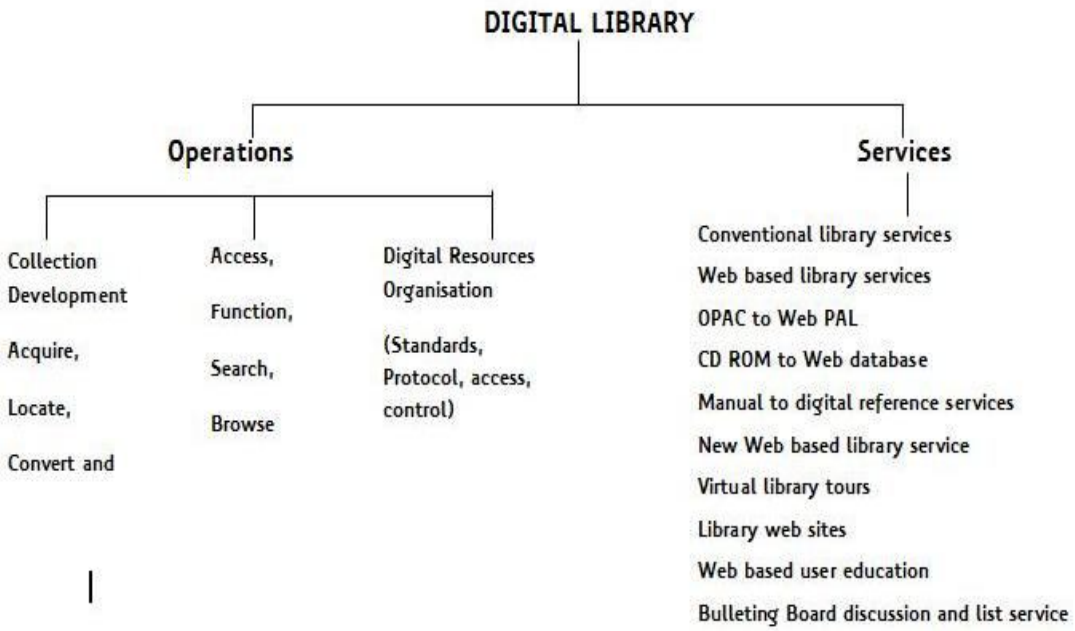
{http://www.emeraldinsight.com/Insight/viewPDF.jsp?contentType=Article&Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2800400404.pdf}

A digital library is much more than just the collection of materials in its depositories. It provides a variety of services to all its users. The basis of digital library is the information objects that provide the content on the form of digital resource.

Thus, the Digital Library is a solution involving to overcome the problems of storage, faster retrieval and archival by application of Digital Computing Machinery, Multi-Level Storage Mechanism, Distributed Multi-Media content and Associated software required to manipulate, extract, interact, emulate, retrieve, collect, catalogue, customize or find any part of this content and selectively add/modify any relevant content.

Misconception: - A digital library is not just another data warehouse. It can be visualized as a type of Information Retrieval System.

Having a sneak preview of Digital Library in a usage context, it is much promising for the end-users and solution to many challenges faced by the system designers / developers.



The collection of a digital library may include a combination of structured or unstructured text, numerical data, scanned images, graphics, audio and video-recordings.

5. Creation of Digital Libraries

5.1 Digital Collection

(i) Converting paper and other media in existing collection to digital form.
(ii) Acquisition of original digital works created by publishers, institutions and other scholars like electronic books, electronic journals and data set.

A digital library can have wide range of resources. It can contain both conventional documents as well as information contained in digital or computer accessible form. Rusbridge has divided resources for a digital library into different categories which are - legacy, transition and new:

- **Legacy:** Legacy resources are largely non-digital resources including manuscripts, prints, slides, maps, audio and video recordings. Attempts are being made to digitize these resources.

- **Transition:** resources are primarily designed for another medium (mostly print). These are being or have been digitized, making the transition into the digital world.

- **New Digital Resources:** New resources are electronic journals, electronic books through database and data sets in many formats. Publishers are increasingly moving to XML or SGML format.

The acquisition of documents which are already available in digital formats like CD-ROM database is also possible. Now-a-days a large number of information products are available on CD-ROM, like MEDLINE, COMPENDEX, METADEX, etc.

5.2 Conversion of Hard Copy of the Documents into Digital Format

The existing selected conventional collection of a library is being converted into digital form with the help of scanners. Printed text, pictures and figures are transformed into computer accessible forms using digital scanner or a digital camera.

There are basically two methods for converting the print resources into the digital resources. These are:

(i) Scanning and use of OCR Programs
(ii) Re-keying of data

Most scanning software generate default TIFF format which can be converted into PDF format using number of softwares. The optical character recognition (OCR) is another method of converting print to digital. The two well known available OCR techniques is that of Xerox Textbridge and Caere's omnipage incorporation technology. Retaining the page layout is the major problem in OCR. Now a number of softwares are available for retaining. The process of retaining the page layout is software dependent. The process of OCR results in computer processible file is that is less accurate than re-keying in the data. The acrobate capture 2.0 provides several options for retaining not only the page layout, but also the fonts, and to fit text into the exact space occupied in the original. In image+ text solution or OCRed text is generated for each image where each page is an exact replica of the original and left untouched.
The re-keying (or data entry) or retyping is another mode of converting printed text into digital form. This involves complete typing of the text. This process although comparatively more accurate, is quite time consuming and costly also. But corrections, modifications and changes in formats are easier to carryout.

5.3 Access to Digital Information Available on Web

These days, web provides the hyper-media based systems that allows rapid access to a wide variety of networked information resources. One can browse the different web sites which are scattered geo-

graphically and have access to the resources. Some of the portal sites or gateways that provide access to electronic resources are.

http://www.edoc.com/
http://mel.library.mi.us
http://bubl.ac.uk
http://sunsite.berkeley.duc/internal index/

So a digital library may have

(i) collection acquired in a digital form
(ii) collection digitized in house
(iii) purchasing access to electronic resources
(iv) gateways to electronic collections available on the web.

6. Digital Library and Its Use

The important functions and use of a digital library in context to users are:

(1) Provides access to very large information collection in a digital form;
(2) Supports multi-media content;
(3) Network accessible;
(4) Provides user friendly interface;
(5) Offers links to local/external objects;
(6) Supports advance search and retrieval;
(7) Supports traditional library mission of collection development, organization, access and presentation;
(8) Supports publishing, annotation and integration of new information;
(9) Brings together people with formal, informal and professional learning missions;
(10) Provides faster access to information resources; and
(11) Provides easy mechanism for resources sharing with other libraries. Sharing of digital files is much easier.

7. Issues and Challenges before Digital Library

(i) Technical Architecture: Libraries need to enhance and upgrade current technical architecture, such as:-

- High speed local network and fast connection to internet,
- Relational database that supports a variety of digital formats,
- Full text search engines to index and provide access to resources,

- A variety of servers such as web services and FTP servers,
- Electronic document management system.

(ii) Building Digital Collections: Libraries will digitize their existing material and acquire original digital works is to be decided first. Who collects and/or digitize which material could be based on such factors as collection strength, unique collections, the priorities of user groups, manageable portions of collection, technological resources and skills of the staff.

(iii) Digitisation: What portion of the collection is to be digitized. There are several approaches available, at least theoretically.

- Retrospective conversion of collections
- Digitisation of particular special collection or a portion of it.
- Highlight a diverse collection
- High use materials
- An adhoc approach – one digitizes and stores material as they are requested.

(iv) Metadata: Metadata is important for digital libraries because it is the key to resources, finding tool for any document like a library catalogue. The "Dublin Core" is one of the prominent schemes. There is problem of naming, identifiers and persistence. There naming is required to uniquely identify digital objects. Any system of naming should be permanent, lasting indefinitely. Three schemes proposed to get over the problems of persistent naming are PURLs, URNs, and Digital Object Identifiers:

(a) PURLS – are persistent URLs a scheme developed by OCLC to separate document name from its location.

(b) URN – Uniform Resource Name have been developed by the Internet Engineering Task Force (IETF).

(c) Digital Object Identifier (DOI) – Developed by the Association of American Publishers and Corpo ration for National Research Initiatives to provide a method by whcich digital object can be reliably identified and accessed.

(v) Copyright/rights Management: Copyright is one of the most important barrier to digital library development. The current paper based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects are less fixed. Easily copied and remotely accessible by multiple users simultaneously. So libraries will ever be able to freely digitize and provide access to the copyrighted material in their collection. They have to develop mechanism for managing copyright.

(vi) Preservation: Another important issue is preservation. In the preservation of digital material, the real issue is technical obsolescence. There are three issues involved in preservation:

- Preservation of the storage medium – tapes, hard drives, floppy discs have a short life span when considered in terms of obsolescence.
- Preservation of access to content – this form of preservation involves preserving access to the contents of the document regardless of their format.
- While files can be moved from one storage medium to another, what happens when the formats containing the information become obsolete.

8. Frameworks (Case study of Stanford Digital Library Architecture)

In general a Digital Library is built around repository software & an information system to provide digital preservation and extract information is developed. Stanford Digital Library is thinking of an innovative Digital Library information handling architecture called Info bus (Information Bus). It basically focuses on Asynchronous information. Primary holding or content of this Digital Library is the Meta Data. Stanford Digital Library, apart from providing typical library services, it also aims to provide value added services like bibliographic generation, citation analysis, billing payment, security & authentication etc.
In principle, content of the Digital Library may be of any form: text, audio, video, graphics, image & 3-D models.

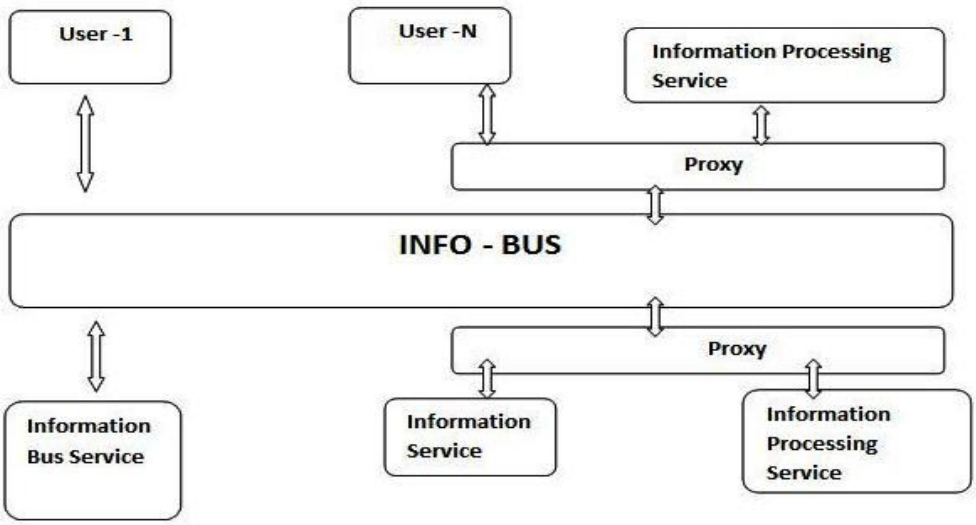


Figure:The Stanford Digital Library Architecture Diagram

A digital library can generally comprise of following:

- Digital Library Resource Manager
- Digital Metadata / Content Manager
- Application Enabler (Optional)
- A multi-agent system comprising of many agents including a monitoring agent
- Digital Library Security Manager

9. Digital Library Breakthrough: - Multifaceted Data Integration & Internetworking Aspects

With the advent & exploration of digital library (much driven by the need of hour), life is easy but there always exists a scope for more to be done in terms of services which eventually becomes need. In the Digital Library context much is done towards text data format but there are glamorous applications which are looking beyond conventional boundaries to support multi faceted data (text + audio + video + image etc...). With the integration of multi faceted data **(Multimedia)** to Digital Library the so called glamorous application like distance learning, virtual library access & living books are not a distant dream. e.g. – Virtual library access means that one instantly has access to all published material in the world in its original form & format (whatever it be text, audio, video etc...) & that can be easily browsed, print and displayed.

10. Benefits in the field of Library Science

- Search criterion got diversified which are :-
  Texture based search
  Shape based search
  Color based search
- Retrieval processing is rejuvenated for audio & video data by streaming process in a streamlined way.
- Storage is done not in raw formats, it's compressed.

So, accessing has become simple, but with Complexity/disadvantages reached next level (nothing comes without a price) for the multimedia processing in terms of:-

- Novel communication & networking technologies for multimedia database.
- Need for truly integrated media system to connect.
- Advance technologies for indexing & retrieval of visual information in large archives.
- Integration of video, audio, text animation, graphics, knowledge about them & their interrelation ship in real time.
- Storage, compression & decompression of the data format without rupturing the content.
- Searching on different parameters for the data format could be very much expensive.

11. Internetworking

The link which serves the purpose is information available over the internet. As of today, internet is the driving factor for communication between the remotely located information. Digital library is no exception to this, there are developments taking place around the world to handle multimedia objects in the distributed environment.

Primarily the Digital Library research concentrates upon Internet-based Distributed Multimedia Information Systems. Organization, storage, Search, Retrieval, Customized Document Services, Intra-document Media Object search & transparency across a variety of platforms are the main focused areas along with asynchronous vs. synchronous services.

With the evolution/discovery of internet for multimedia communication, it paved a path for personalization setup of distributed digital library in the world of ever rapidly increasing information (in a distributed way). Evolving network infrastructure enabled large information repositories to be accessed virtually from anywhere. These trends have a great impact for researchers/users to work effectively by being prepared to look for digital sources (including the information available over network). E.g. preparation an environmental impact report might need to gather information from several sources, interacting with a geographical information server at one location to obtain maps & related geographical data & an image server to fetch satellite imagery report. But in the Digital Library, is a variety of personalization of information space is an important capability.

New digital processes will characterize future information systems (e.g. agents, user, profiling etc) would require even further personalization & customization functionality than available in contrast to distributed support.

12. Applications of Hypermedia for Digital Library

Hypermedia application or systems are the retrieval application, which provide flexible access to multimedia information and a novel method to create structure & manage database in a point-to-point or multi-point fashion. It provides the user more freedom in assimilating and exploring information. Taking the case over internet, for technical papers / reports that are considered more suitable for linear reading in sequence for which HTML (Hypertext Markup Language) is included. Hypertext is used for text only & Hypermedia is used for multimedia data.

The system requires user interfaces, powerful authoring & presentation tools. There are couple of applications which are particularly suited for hypermedia such as encyclopedia, dictionaries & other reference

books. These are comprised of a number of independent units that are seldom accessed sequentially but rather by selection & cross-reference with other entries.

13. Features of Hypermedia Systems:-

- Information is divided into smaller units or nodes which contain single media data such as text, audio, video, graphics or animation etc...
- The links that interconnect the units of information can be bidirectional to facilitate backward traversals.
- Nodes or Links are distributed across multiple sites & remote servers in a computer network.
- Linear reading is for document & Non sequential browsing to traverse in information space and navigate hypermedia network.
- Own information structure can be created & shared.
- Maintain a DBMS to manage the shared multimedia data.

14. Impact of Hypermedia on Digital Library

With the overview of what is hypermedia data & application we are sound enough to proceed how these make a significant impact on the Digital Library. Many of the advantages of Digital Library are immediate consequences of using hypermedia technology. Hypermedia functionalities such as navigation & annotation are essential in a Digital Library context, allowing the generation of multiple multifaceted relationships inter and intra documents. The hypertext dimension of the framework allows navigation via static & dynamic hyperlinks.

Implemented typical examples of Hypermedia in a Digital Library context

- Extraction of multimedia data from a Digital Library via a response to query which are called bundles.
- Construction of geographical digital libraries using a hypermedia framework.
- Development of Digital Library based hypermedia infrastructure to support scholarly work in distributed environment.

15. Implementation Issues of hypermedia in Digital Library

- Interoperability: - Due to proprietary storage mechanism & document formats; most current hypermedia systems are amenable to interoperability.
- Integration: - It is difficult to integrate the materials created in different systems.

- Architecture: - There are much architecture discussion floating for hypermedia in Digital Library but there isn't as such concrete one.
- Searching: - The traversing for search from linear to non linear in a distributed environment is having implementation issues.

16. Conclusion

The main objective of the reengineering process was to create flexible platforms with open interfaces within the overall organizational structure of the library, in order to meet the challenges of the digital world. The modern technology, like – hypermedia, multimedia, digitization of conventional documents, etc. has changed the scenario and work culture and atmosphere of library as the Library staff has also become more efficient to compete with the growing era of modern technology. Likewise, users are also becoming more high-tech and they desire the data in digital or electronic form. Thus, as the time passes, the library world would become digital and then will lead to paperless library.

References

1. HILDEGARD SCHAFFLER, "How to organise the digital library: reengineering and change management in the Bayerische Staatsbibliothek, Munich" http://www.emeraldinsight.com/Insight/viewContentItem.do;jsessionid=5A7A7DD6EA88A3CACD2BFA510624E51F?contentType=Article&contentId=861437
2. MARCOS ANDR´E GONC¸ALVES "Constructing Geographic Digital Libraries using a Hypermedia Framework" Multimedia Tools and Applications 8, 341–356 (1999)
3. BOOK: PRRASHER, R.G. (Ed.). "Indian Libraries in IT Environment"; edition 2003 Medallion Press
4. BHARDWAJ, RAJESH KR. and SHUKLA, R.K., "Re-Engineering of Library and Information Services through Web Modeling at Delhi College of Engineering".

About Authors

Mr. Ram Jha, Software Engg., J P Morgan, Mumbai,
E-mail:ramsjha@gmail.com

Mr. Abhishek Kumar, Scientist - B, INFLIBNET Centre, Ahmedabad
E-mail: abhishek@inflibnet.ac.in