
DIGITAL LIBRARY ARCHITECTURE: A CASE STUDY

Sewali Bhuyan

Abstract

The time for comparisons and conflicts between print media and electronic media appear to be over as the world prepare itself for the next era in the digital society. The need of modern information society has changed the library from traditional type to digital one. This paper discusses problems challenged and issues involved in design and development of worlds standard digital library focused on India Context. The paper explains the concept and architecture of a digital library. Digital libraries can immediately adopt innovations in technology providing users with improvements in electronic and audio book technology as well as presenting new form of communication. This Paper also discusses the Metadata concept as it is developed in the field of library and information science. If is an effective way of information management. It briefly touches upon concept of Metadata repository. It also discusses how Indian University Libraries can develop a digital library for electronic theses and Dissertation (EID). Finally, it highlights the initiative taken by various Indian Organization towards the development of Digital Libray.

Keywords : Digital Library, Metadata, EID

1. Introduction

The concept of library is going through a revolutionary phase due to the proliferation of electronic resources. Our methods of producing, organizing and seeking information have changed drastically with the usage of computers and databases, but our problems have not been solved.

The library professional had never been exposed in the past to the changing information scenario as they are being exposed now. Information explosion and the development of technologies and its progress is changing the previous methods of document collection, storage and dissemination. Now librarian have to face the challenge of this changing scene, otherwise they can be replaced by those, who are able to disseminate the information through CD networks, digital libraries, electronic publishing and Internet etc. So the librarians well have to fulfill this obligation as well.

The level of interest regarding digital libraries has grown steadily as a greater number of institutions, including archives and Museums consider the possible implication of digital libraries while there are important unresolved digital library research and development issue, there is also a concurrent desire to develop strategies for systematic digital library programs built upon the result of digital library Project.

2. Digital Library Concept

Digital library programs generally include both digital collections and services that facilitate access, retrieval and the analysis of the collection. A digital library is a collection of information that is both electronic as well as digitized and it gives us power we never had with traditional libraries. The term digital libraries were first popularized by the NSF/DARPA/NASA Digital Libraries Initiative in 1994. A digital library is a library in which collections are stored in digital formats and accessible by computers. In the Kahn/Wilensky architecture, items in the digital library are called "digital objects". They are stored in "repositories" and identified by "handles". Information stored in a digital object is called "content" which is divided into "data" and information about the data, known as "properties" or "metadata".

2.1 Key Component

Fully developed digital library environment involves the following elements.

- 1) Initial conversion of content from physical to digital form
- 2) The extraction or creation of metadata or indexing information describing the content to facilitate searching and discovery as well as administrative structural metadata to assist in object viewing, management and preservation.
- 3) Storage of digital content and metadata in an appropriate multimedia repository. The repository will include rights management capabilities to enforce intellectual property rights, if required E-commerce functionality may also be present if needed to handle accounting and billing.
- 4) Client services for the browser, including repository querying and workflow.
- 5) Content delivery via file transfer or streaming media.
- 6) Patron access through a browser or dedicated client
- 7) A private or public network

These components might not all be part of a discrete digital library system but could be provided by other related or multipurpose system or environment. Accordingly, integration is a consistent issue cited by digital library developers.

2.2 What is Metadata

Metadata describe the attributes and contents of an original document or work and describes a resource. Metadata may be defined as representing higher-level information that describes the content, context, quality, structure and accessibility of specific data sets such as digital data images, databases and printed materials. As large scientific databases were developed, it became evident that surrogates were required to provide more information about data sets:

Typically metadata include two types of information

- 1 Basic details about the institutions that hold relevant information who are they? where are they and what is their function? What are their.
 - Available resources?
 - Key linkages (who is currently working with whom and how)?

- 2 About relevant data sets :
 - Description of data sets (What, purpose, form at and how managed) .
 - Coverage (geographic, thematic, time scale, completeness, limitations and gaps) ,
access (availability, cost, formats available and documentation) .

Metadata not only provides pointers to the original data sets but it also help in sharing data among the database produces. It is a tool to integrate data that are in heterogeneous format and scattered geographically, several agencies are taking initiatives in creating Metadata / Metadata base by using various Metadata standard.

3. Architecture of Digital Library

The architecture of the digital library as describe by Kahn and Wilensky, specifies those characteristics that apply to all types of material. To example object needs to save a name or identifier. Names are a vital building block for the digital library. Names are needed to identify digital objected, to register intellectual property in digital objects, and to record changes of ownership. They are required for citation for information retrieval and are used for links between objects. These names must be unique. This requires a administrative system to decide who can assign them and change the objects that they identify. They must last for very long time periods, which exclude the use of an identifier tied to a specific location, such as the name of a computer. Names must persist even if the organization that named an object no longer exists when the objects is used. There need to be computer systems to resolve the name rapidly, by providing the location where an object with a given name is stored.

The corporation for National Research Initiatives has implemented a handle system which satisfies these requirements. A "handle" is a unique string used to identify digital objects. The handle is independent of the location where the digital object is stored and can remain valid over very long periods of time. A global server provides a definitive resource for legal and archival purpose, with a caching server for fast resolution. The computer system checks that new names are indeed unique, and supports standard user interfaces, such as Magic. A local handle servers is being added for increased local control.

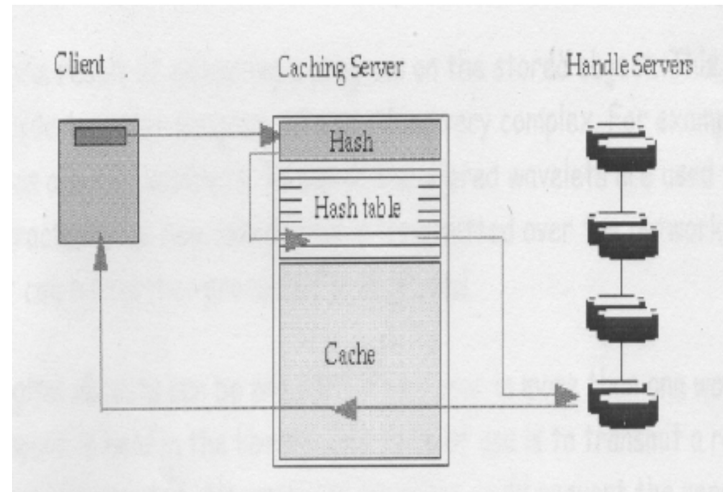


Figure: 1. The CNRI handle system

3.1 Digital Library Objects are more than Collection of Bits

In the digital library, information is stored as "digital objects". A primitive idea of a digital object is that it is just a set of bits, but this idea is too simple. The content of even the of the basic digital object has some structure, and information, such as intellectual property rights, must be associated with the digital object. Figure 2 shows that object in a repository has two parts, content and associated data, sometimes called "metadata".

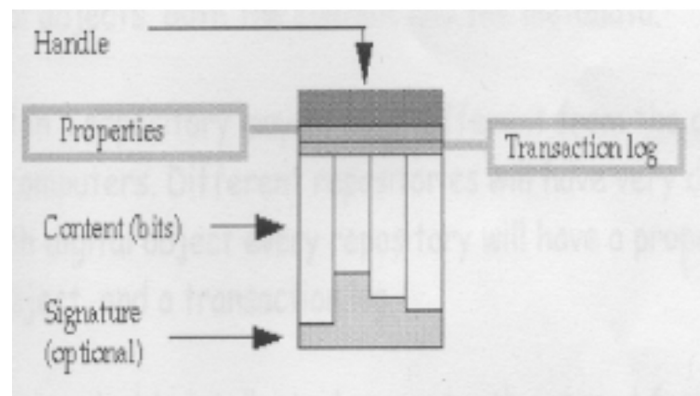


Figure : 2. Parts of a digital object

3.2 Repository

A repository stores digital objects, both the content on the metadata. A digital objects as stored in a repository may be very different from the digital object that is made available to users' computers.

Different repositories will have very different internal organizations, but for each digital object every repository will have a properties record, which holds attributes of the object, and a transaction log. Since digital objects contain valuable intellectual property, the stored form of a digital object within the repository includes information that allows for it to be managed within economic and social frameworks. The repository maintains this information, provides basic reference information, and provides security to ensure that only valid operations are carried out on the digital objects.

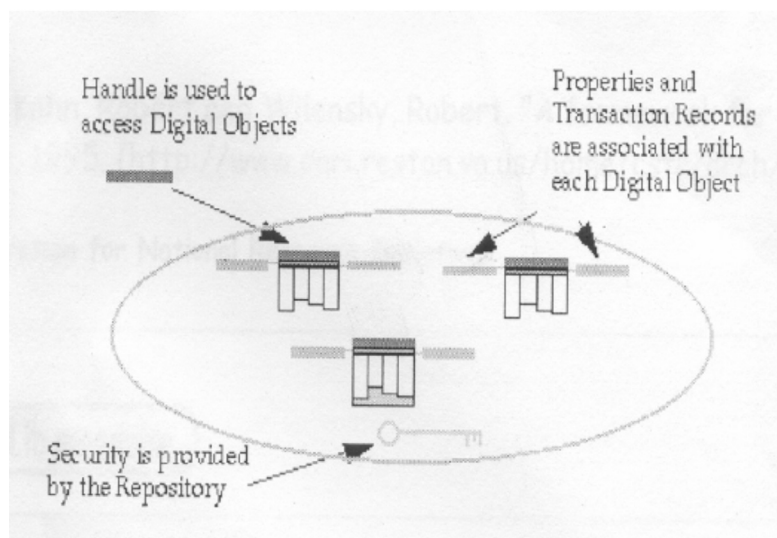


Figure: 3 A repository

The internal organization of a repository and the way that digital objects are stored are hidden from the user. A simple protocol is called the "repository access protocol." The basic commands in this protocol are those to access a digital object and its metadata, and the service request to disseminate a digital object. In addition there are commands to add and delete digital objects.

4. National Science Digital Library

National Science Digital Library (NSDL) is a wide-ranging program of the National Science Foundation (NSF) to build library collections and services for all aspect of science education.

The Metadata Repository

The Metadata Repository is a key component of NSDL architecture. Its function is to support providers of services, such as the NSDL search service; its function is support providers of services, such as the NSDL Search service. It holds collection level Metadata about every collection known to the NSDL and an item level metadata record for each known individual item.

There are two reasons for exporting metadata. The first reason is to supply metadata to service providers who are contributors to the NSDL, such as the NSDL search service. The metadata repository provides a single location from which such providers can obtain metadata about all NSDL collection and the resource within them. The philosophy of NSDL is to in coverage highly creative uses of the available resources.

The second reason for exporting metadata from the repository is to encourage other digital library developers to consider the NSDL as a resource that they can build upon. In building a digital library, every organization has a tendency to envisage architecture with its own library at the center.

5. Digital Library Services

Digital library services provide consulting; applications and information to help libraries, create manage and preserve digital content for the use of the university system. Digital library service is helping the state university libraries put digital collection online.

- PALMM – Publication of Archival, Library museum.
- Support for electronic theses and Dissertation (EID) .
- 'New' state university system ETD workshop.
- Cost effective central storage facilities.
- Tools for loading, storing, searching and displaying digital projects.

5.1 Electronic Theses and Dissertation

Electronic theses and dissertations (EID) are defined as those theses and dissertations submitted archived, or accessed primarily in electronic formats. That includes additional word processed (a typusillen and scanned) documents made available in PDF, as well as less traditional hypertext and multimedia formats purchased electronically on CD – ROM an World Wide Web.

There are many reasons for having EID

- Almost all TD's are produces as electronic documents and if researchers know in advance about have to prepare EID, then creating their own EID usually is very simple process.
- Minimize duplication of effort.
- Improve visibility.
- Accelerate EID s available faster to outside audience.
- Cost and benefits.
- Enhancing access to university research.
- Helping universities develop digital library services & infrastructure.
- Increasing sharing collaboration among universities and students.

Traditional methods of archiving and storing theses and dissertations are inefficient and unwieldy. Many theses and dissertations lie mouldering in library stacks, with no efficient way for researchers to locate the information that may be contained in them. Further the time and cost involved in procuring copies of those works may often be prohibitive. The main objectives are as follows :-

- To improve education and research by allowing students to produce electronic documents, are digital library and understand issues in e-publishing.
- To lower the cost of submitting and handling theses and dissertations.
- To empower students to convey a richer message through the use of multimedia and hypermedia technologies.
- To empower universities to unlock their information resources.
- To advance digital library technology MDLT.

Technical issue involved

Tools for creation, management, access, archiving and storage:

Metadata:

- What information regarding ETD can collect and share.
- Dublin core and resource description format.
- XML and ETD metadata (ETD – MS: an interoperability metadata standard for electronic theses and dissertations).
- VILS union metadata service for NDLTD format for ETD.
- Document format of ETD (PDF or XML)
- Capable of complete full text retrieval.
- The information retrieval engine.
- Copyright and publication multilingual system.

5.1.1 ETD in India

Indian universities play a major role in generation and dissemination of knowledge by conducting research works and producing PhD theses as a unique genre of information sources. As per reliable sources, every year, an average of 8000/- 100000 PhD are awarded in India.

UGC Infonet, an ambitious programme of UGC is around and university libraries have to best utilize it for content creation and management. As part of ongoing international efforts to networked digital library of theses and dissertations Indian university libraries can also develop a digital electronic theses and dissertations (ETDs). This can be done in two ways. While ETDs are owned and maintained by the institutions at which they were produced or archived, it is possible to give searchers the appearance of a single collection by gathering all the metadata (title, author etc.) into a central search engine. Then when a potentially relevant document is found, the user will be redirected to

the institution that contains the actual document. Otherwise theses in e-form can be sent to INFLIBNET, where we can host them, and allow users to browse through and download them. INTLIBNET has already hosted an online database of theses containing around 1.4 lakhs of bibliographic records of PhD submitted to Indian universities. Full text of existing theses collection can also be made available by converting them into digital form.

6. Digital Library Development issues in India

There are uncounted numbers of problems the Digital library development teams face in India while they embark on the digital development as well as during progress phase. Some of the prominent and predominant among them include the following.

(i) Lack of Proper Information and Communication Technology (ICT) Infrastructure. Digital Libraries Demand Cutting Edge IT and Communication Infrastructure such as

- (a) High end and powerful server; structure LAN with Broadband Intranet facilities ideally optical fibre based Gigabit networks;
- (b) Required number of workstations capable of providing online information services, computing and multimedia application.
- (c) Internet connectivity with sufficient bandwidth, capable of meeting the informational and computational requirement of the user community.
- (d) Lack of proper planning and Integration of Information resources: presently the library acquisitions in India are either paper based and electronic. Some of the libraries need retro-conversion and digitization of library holding too. Literature on related studies show that there is a severe lapse on the libraries with regard to proper planning of the Information resources which are conducive for developing digital libraries.

There is a dire need for proper planning and meticulously framed content integration model which is achieved and implemented through world standard digital library technologies.

(ii) Rigidity in the Publisher's Policies and Data Formats.

Having successfully installed and configured a digital library does not qualify a library to automatically populate all its digital collection into the digital library. One has to obtain publisher's consent and copy right.

Permissions for the same digital libraries software usually accept and process all popular and standard digital formats such as HTML, word, RTF, PPT, or PDF. Most of the publisher's put their materials in their own proprietary e-book reader formats, from which the text extraction became almost impossible.

(iii) Lack of TTC Strategies and Policies

A vast majority of the libraries in India do not have laid down policies on TTC planning and strategies to meet the challenges posed by the technology push the information overload, as well as the demand pull from user.

(iv) Lack of Technical Skill

The Human Resource available in the libraries need time to time professional enrichment inputs and rigorous training on the latest technology which are playing around in the new information environment. The kind of training programmes being imparted in India at the moment are not able to meet the demand in terms of quantity as well quality.

(v) Management Support

For the provision of world class Information system, resources and services the libraries need the wholehearted sport from the respective management. Institutional support in terms of proper funding, human resources and IT skill enrichment are pre-requisites for the development and maintenance of state-of art digital library system and services.

(vi) Copyright/ IPR Issues

Issue of Copyright, intellectual property and fair use concerns are posing unprecedented array of problems to the libraries and librarians are struggling to cope with all these related issues in the new digital environment

6. Conclusion

Development of digital library needs the availability of appropriate infrastructure tools and techniques and also manpower. In India concept of a digital library is new phenomenon and we do not have efficient library experts in the country who are also well trained in these digitizing process. There is a need for adequate number of highly trained staff because converting library into digital one is complex task. In India there have been institutions like INSDOC, INFLIBNET arranging training for library professionals in the use of digital resources and development of a digital library in the networking environment.

Department of library and Information science in the universities have been providing some basic training in library automation which indeed has not been sufficient at all for equipping library professionals for handling library automation job. It is heartening to note that Greenstone, D space and E-print installation are picking up quite fast in India and institution like DRIC, INFLIBNET ,

NCSI, IITs, IIMK and many other are giving wide popularity and training on these software. India has recognized the power of digital libraries and lots of initiatives are on the move for developing a digital library.

References

- (1) Arms, William Y, Dushay, Naomi; Fulker Dave and Lagaze Carl. A case study in metadata Harvesting : the NSDL
- (2) Arms , Willam Y. Key Concept in the Architecture of the Digital Library. D-Lib Magazine, July/1995.
- (3) Digital Library .From Wikipedia
- (4) E-Resource Management using UGC Infonet. January 05-09. 2004. [http:// www.inflibnet.ac.in](http://www.inflibnet.ac.in)
- (5) Nagarkar, Shubhada.. Metadata : A data Integration tool. XXII IASLIC conference, 1999.
- (6) Sreekumar, MG and Sreejaya, P. Digital library initiatives and issues in India : efforts on scholarly Knowledge Management. India Institute of Management Kozhikode.

ABOUT AUTHOR

Mr. Mahendra Maheta is Chief Librarian at Vivekananda Institute of Hotel and Tourism Management, Rajkot, Gujarat.

E-mail: librarian007@rediffmail.com