Somvir, Sudha Kaushik / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 1, Issue 2, pp.086-091
www.ijera.com

THE ARCHITECTURE OF DATA STORAGE SYSTEM FOR A DIGITAL LIBRARY

Somvir
Research Scholar
Dept. of Library & Information Sc.
Singhaniya University, Rajasthan

Sudha Kaushik
Librarian
PDM College of Engineering, Bahadurgarh, Haryana

ABSTRACT:
Today's society is the information society and the libraries are moving from traditional towards digitalization. The digitalization is the result of the need of modern information society. The paper explains the concept and architecture of a digital library. Problems, challenges and issues involved in design and development of standard digital library focused on Developing Countries Context also discussed in the paper. Digital libraries can immediately adopt innovations in technology with improvements in electronic and audio book technology, presenting new form of communication. The authors also discuss the Metadata concept as it is developed in the field of library and information science. The concept of Metadata repository briefly described. It also discusses how we can develop a digital library for electronic theses and Dissertation.

Keywords: Digital Library, Metadata, Handle System, Repository, Electronic Theses & Dissertation

INTRODUCTION:
The emergence of digital technology and computer networking have provided means whereby information can be stored, retrieved, disseminated, and duplicated in a very fast manner to meet the context, the right information to the right user at right time. Digital libraries have made considerable advances, both in technology and its applications. The digital library initiatives at international level are many, but in developing countries they are still in a nascent stage. With the use of computers, latest IT and databases, our methods of producing, organizing and seeking information have changed drastically. In present the library & information professionals are being exposed to changing the information scenario, they had never been exposed in the past. The previous methods of information/ document collection, storage and dissemination are changing by the information explosion & the development of technologies and its progress. Now the traditional librarians have to accept the challenge of this changing scene and work according the conditions, 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.

DIGITAL LIBRARY:
The collection of information in electronic & digitized form is the base of digital library and it gives as power we never had with traditional libraries. The digital library federation in the USA defines the digital library as: Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities. A digital library is an organized collection of digitized material or it's holding in the digital form, which can be accessible by a computer on the network by using TCP/IP or other protocol. Digital collections and services that facilitate access, retrieval and the analysis of the collection are generally included in the Digital library programs. The term digital libraries were first popularized by the NSF/DARPA/NASA Digital libraries initiative in 1994. 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".

The Digital Library is:

1. Organized collection of multimedia and other types of resources.
2. Resources are available in computer process able form.
3. The function of acquisition, storage, preservation, retrieval is carried out through the use of digital technology.
4. Access to the entire collection is globally available directly or indirectly across a network.
5. Support o users in dealing with information objects
6. Helps in the organization and presentation of the above objects via electronic/digital means etc.

Elements of the digital library

Fully developed digital library environment involves the below mentioned elements. These components might not be 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.

- A private or public network
- Client services for the browser, including repository querying and workflow.
- Content delivery via file transfer or streaming media.
- Initial conversion of content from physical to digital form
- Patron access through a browser or dedicated client
- Storage of digital content and metadata in an appropriate multimedia repository, including right management capabilities to enforce intellectual property rights, if required E-commerce functionality may also be present if needed to handle accounting and billing.
- 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.

METADATA

The term refers to any data used to aid the identification, description and location of networked electronic resources. Many different metadata formats exist, some quite simple in their description, others quite complex and rich. Metadata is defined as data providing information about one or more aspects of the data, such as:

Means of creation of the data; Purpose of the data; Time and date of creation; Creator or author of data; Placement on a computer network where the data was created and the standards used

The metadata of a text document contains the information about the length, author, time of written and summary of the document. And in case of digital image, metadata describes how large the picture is, the color depth, the resolution, when it was created, and other data. Metadata is data. As such, metadata can be stored and managed in a database, often called a registry or repository. However, it is impossible to identify metadata just by looking at it because a user would not know when data is metadata or just data.

METADATA IN LIBRARIES

Metadata has been used in various forms as a means of cataloging archived information. The DDC system employed by libraries for the classification of library materials is an early example of metadata usage. Library catalogues used 3x5 inch cards to display a book's title, author, subject matter, and a brief plot synopsis along with an abbreviated alpha numeric identification system which indicated the physical location of the book within the library's shelves. Such data helps classify, aggregate, identify, and locate a particular book. Another form of older metadata collection is the use by US Census Bureau of what is known as the "Long Form." The Long Form asks questions that are used to create demographic data to create patterns and to find patterns of distribution. The term was coined in 1968 by Philip Bagley, one of the pioneers of computerized document retrieval. Since then the fields of information management, information science, information technology, librarianship and GIS have widely adopted the term. In these fields the word metadata is defined as "data about data". While this is the generally accepted definition, various disciplines have adopted their own more specific explanation and uses of the term.

Metadata describe the attributes and contents of on 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 set such as digital data images, databases and printed materials. As large scientific databases were developed, it become evident that surrogates were required to provide more information about data set:

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.

ARCHITECTURE OF DIGITAL LIBRARY

Kahn and Wilensky describe the architecture of the digital library having the characteristics that can apply for all type of material. A name or identifier is essential to save and object. For the digital library the names or identifiers are a vital building block, which are needed to identify digital objected, to register intellectual property in digital objects, to record changes of ownership, required for citation for information retrieval and are used for links between objects. These names/ identify must be unique. An administrative system is required 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 and the names must persist even if the organization that named an object no longer exists when the objects is used. The computer systems are required to resolve the name rapidly, by providing the location where an object with a given name is stored. To achieve these satisfactions a handle system is implemented. A "handle" is a unique string used to identify digital objects and it 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 server is being added for increased local control.

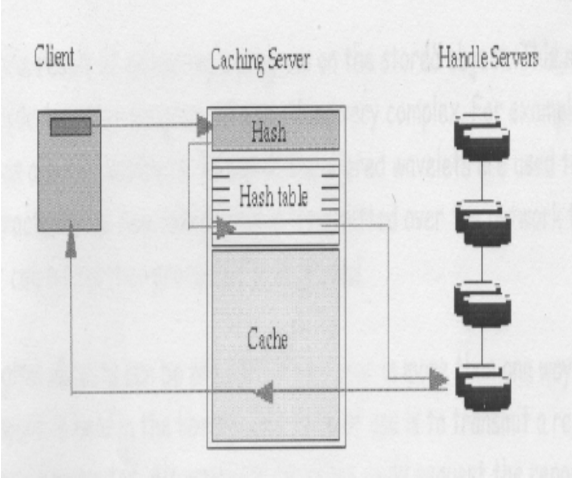

Figure: 1. The Handle System

Parts of Digital Library Objects

Information is stored as "digital objects" in the digital library. 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".

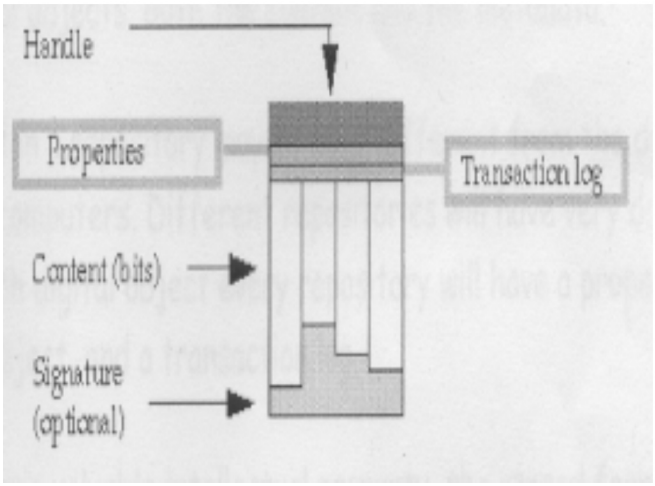

Figure : 2. Parts of a Digital Object

REPOSITORY

A repository stores digital objects, both the content on the metadata. A digital object 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 from 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.

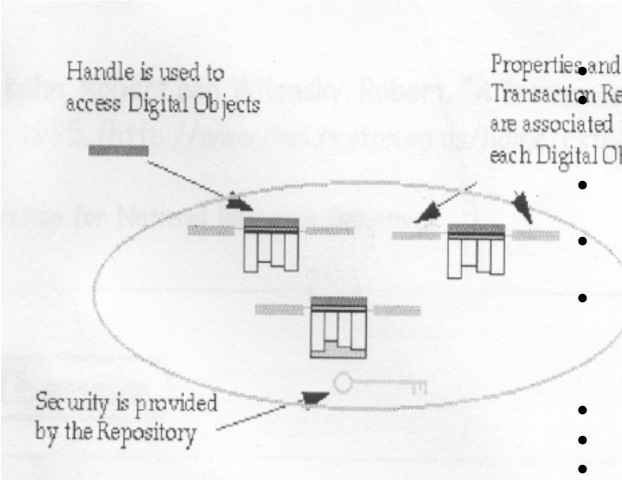

Figure: 3 A Repository

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.

ELECTRONIC THESES AND DISSERTATION

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

Needs of ETD

- Almost all TD's are produces as electronic documents and if researchers know in advance about have to prepare ETD, then creating their own ETD usually is very simple process.
- Minimize duplication of effort.
- Improve visibility.
- Accelerate ETD 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.

Objectives for ETD

The traditional methods of archiving and storing theses and dissertation are inefficient and unwieldy. Many theses and dissertation lie moldering 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 advance digital library technology.
- 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 improve education and research by allowing students to produce electronic documents.
- To lower the cost of submitting and handling theses and dissertations.

Technical issue involved

- Tools for creation
- Management
- Access
- Archiving and storage

Metadata:

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

ETD in India

Through conducting research works and producing PhD theses as a unique source of information, Indian universities play a major role in generation and dissemination of knowledge. UGC INFONET, on ambitious programmed of UGC is around and university libraries can do best utilize it for content Creation and management. As part of ongoing international effects to networked digital library of theses and dissertation Indian university libraries can also develop a digital electronic theses and dissertation (ETDs). Fifteen universities registered and started contribution of ETD at UGC INFONET Sodhganga. White ETD are owned and maintained by the institutions at which they were produced on 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 be 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 and online database of theses of PhD submitted to Indian universities. Full

text of existing theses collection can also be made available by converting them in to digital form.

ISSUES IN DIGITAL LIBRARY DEVELOPMENT

There are umpteen 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 ICT Infrastructure

Digital Libraries Demand Cutting Edge IT and Communication Infrastructure such as

- High end and powerful server; structure LAN with Broadband Intranet facilities ideally optical fiber based Gigabit networks;
- Required number of workstations capable of providing online information services, computing and multimedia application.
- Internet connectivity with sufficient bandwidth, capable of meeting the informational and computational requirement of the user community.
- 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 become almost impossible.

(iii) Lack of ITC Strategies and Policies

A vast majority of the libraries in India do not have laid down policies on ITC 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 technologies 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 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

CONCLUSION

Appropriate infrastructure tools, techniques and manpower is the basic needs for the development of digital library. Concept of a digital library is new phenomenon in the developing countries and there is a lack of efficient library experts who are also well trained in the digitizing process. The converting task of traditional library into digital library is very complex and for it there is a strong need for adequate number of highly trained staff for better performance. In India, training for library professionals in the use of digital resources and development of a digital library in the networking environment is giving the different institutes INSDOC, NISCAIR, INFLIBNET, DELNET, etc. in different universities all over the country the Department of library and Information science 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. Greenstone, D space and E-print installation are picking up quite fast in India and institution like DRTC, 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 http://en.wikipedia.org/wiki/Digitallibrary

4. E-Resource Management using UGC INFONET. January 05-09. 2004. http://www.inflibnet.ac.in
5. Nagarkar, Shubhada.. Metadata : A data Integration tool. XXII IASLIC conference, 1999.
6. http://archive.ifla.org/II/metadata.htm
7. http://en.wikipedia.org
8. http://www.techterms.com
9. Solntseff N. and Yezerski A. A survey of extensible programming languages. Annual Review in Automatic Programming, Vol 7 No 5. 1974, pp. 267-307
10. http://shodhganga.in