
Managing the Digital Library in the Changing Environment

Shreyasi Parikh

Shweta Pandya

Abstract

Recently the idea of the digital library has moved to the forefront of discussion and research. Digital libraries are those libraries, which have number of machine-readable publications and facilities for remote access to several databases. Their aim is to provide instant access to digitized information and consists of variety of information sources-paper to paperless forms including multimedia. The purpose of this paper is to describe how the information professionals can manage the digital library in the changing environment. It describes how an integrated approach of information technology, infrastructure and human resource management is useful in supporting the fruitful utilization of digital library by the users and enhances their satisfaction. Finally, it highlights the issues of ethical concerns in the digital environment.

Keywords : Digital Library; Infrastructure; Standards; Digital Library Management

0. Introduction

Days have come when the quality of a library is no longer determined by the size of the printed collection and size of the manpower. But, the immense growth in the quantity and cost of information has changed the concept. Now the level of user satisfaction judges the quality. This has led to the new patterns of preservation of information available on network and instant access to the information cutting geographical and time barriers. In addition to large quantity of print media, voluminous data is available in electronic format on online as well as off-line. Thus, the encounter with new formats, new media, new information and new communication channels needs restructuring of present day library facilities and services and this has given a birth to the concept of digital library.

1. Digital Library

The digital library has material stored in a computer system in a form that allows it to be manipulated (e.g. For improved retrieval) and delivered (e.g. As a sound file for playing on a computer) in ways that the conventional version of the material cannot be.

An automated library is not necessarily a digital library. Automation does not make it 'digital'. However, digital library must be automated in some of its essential functions.

1.1 Purpose

The main purpose of digital library is to facilitate access to electronic information, print materials and library services to ensure that the information needs of user community are met, regardless of the location. Digital libraries as global, multilingual repositories of data, knowledge, sound and images invite people everywhere to become users and learners.

1.2 Objectives

Some important objectives of digital library are as follows :

- To add to the quality preservation of library material
- To provide multiple referencing at a time
- To facilitate the wide area usage
- To restore archival material
- To provide security to rare/valuable material
- To introduce new services and/or material
- To Promote collections
- To generate income
- To raise the library profile

1.3 Digital material

Digital material includes both – products and services. Computer readability is an important criteria for digital material. Digitization has two stages, first is the conversion of the Physical medium into digital representation and the second stage is to extract the information from the digitized image. Digital materials are items that are stored, processed and transferred via digital (binary) devices and networks. Digital services are services (such as reference assistance) that are delivered digitally over computer networks.

2. Requirements for a Digital Library

There are various requirements that are essential for planning and implementing the process of the digital library, such as support of the management, staff, equipments, customers/clients and digital content acquisition.

2.1 Budgetary and management support

No redesigning or transformation of the library is possible without the support of management. Transformations in initial stage are always expensive. As time passes, they prove cost-effective. Librarian requires to inform and convince the authorities by first giving them a brief idea about the actual requirement and the project proposal should be submitted to management with solid arguments and facts. Once the work starts, the authorities should be continuously informed about the progress and further requirements (budgetary and others). Any problems in the process should be discussed from time to time and solutions be found out at the earliest to prevent unnecessary delay of the project. Convincing the authorities, getting the project sanctioned and getting required budgetary help in time from the management is the crucial task and needs tactfulness and sufficient knowledge on the part of the librarian.

2.2 Acquisition of digital content

Easily available content resources for the digital library are materials such as indexing and abstracting services linked to full text electronic journals, multimedia presentations and digital copies of publications originally published in print etc. The range as well as availability of digital content are increasing very fast and days are not far when more and more electronic content will be available including full test resources on a range of disciplines from a variety of publishers.

2.3 Infrastructure support

The infrastructure includes necessary hardware, software, communication channels and other physical equipments. For the successful implementation of the project the hardware and the software should be

compatible with each other. Other telecommunication equipments should be supportive to the selected hardware and software, which in the long run helps in providing strong security and reliability to the system and hence improve the overall performance of the digital library services. Selection and development of the infrastructure is a time consuming project and cannot be done overnight. The emphasis of the library system must be on the required minimum level of hardware and software for the users to correctly and efficiently interact with the digital library. Support of the management is crucial at this stage.

2.3.1 Hardware Requirements

2.3.1.1 Computer systems

There are two components to any modern distributed client/server system:

PCs (Stand-alone/client machine) or Network computers

Stand-alone computers have been replaced by computers in networks for most of the processing task. The use of multiple computers linked by a communication network for processing is called distributed processing and widely used form of distributed processing is client-server computing. The advantages of the PCs are that they are very familiar to the users and the major disadvantage is that they should be individually upgraded.

A Network is a group of computers connected to each other, for communicating and sharing resources. The advantage of network computer is that it is simple and do not have peripherals. It can utilize the processing power of the server and automatically upgraded every time it is run.

Server

Server is a computer or program that provides services to other computer programs. Servers are basically specialized into three classes: Database servers with high speed disks and very fast local communications, Application servers with fast processors, and communication servers with fast communications peripherals.

2.3.1.2 Capturing Devices

The following are the main capturing devices useful for developing digital library.

Text : Keyboard, Mouse, Bar-code reader, OCR etc.

Images : Digital camera, Video scanner. Slide scanner, Sheet scanner, Flatbed scanner, Drum scanner etc.

Video : Video cards

Audio : Audio cards

2.3.1.3 Storage Devices

Following tables give an idea as to how much space a character or an image occupies in the system and depending on the requirements, the storage capacity can be added to the system.

Text#

Item	Size	Comments
Characters/page	2000	
Characters/article	10,000	5 pages/articles
Characters/collection	1,000,000,000	100,000 articles/collection
Raw data bytes	1,000 MB	1 byte/character
DB structure overhead	200 MB	1 page/article = 2 KB/article
Index overhead	1,000 MB	100 % of raw data
Bib records overhead	150 MB	500 bytes +200%/article
Subtotal	2,350 MB	
Processing etc.	780 MB	33%
Total	3,000 MB	= 3 GB

Images#

Small collection size of 10,000 images captured at 640 x 480 in 256 colours, with bibliographic record, without commentary or description for each, indexed for structural searches will need:

Item	Size	Comments
Bytes/Image	300,000 Bytes	
Raw data bytes	30,000 MB	100,000 images
DB structure overhead	100 MB	1 KB/image
Feature index overhead	20 MB	200 bytes/image
Bib records overhead	150 MB	500 bytes+ 200%/image
Subtotal	30,370 MB	
Processing raid etc.	10,115 MB	33%
Total		40,460 MB = 40GB

Compression of raw image data can be reduce this to about 20 GB

Audio#

Small collection size of 100,000 audio recordings, half sound (8-bit, 11 kHz & mono) and half music (16-bit, 44 kHz & stereo) of 10 minutes, with a bibliographic record, without commentary and description for each, and indexed for structural searches will need:

Item	Size	Comments
Bytes/Sound clip	600 KB	10 min @ 1 KB/sec
Bytes/music clip	10,200 KB	10 min @ 1 KB/sec
Raw data types	540,000 MB	50,000 sound & sound music
Bib records overhead	150 MB	500 bytes + 200%/article
Subtotal	540,150 MB	
Processing raid etc.	180,000 MB	33%
Total	720,150 MB	= 720GB

Compression can reduce the raw data size by about half (360 GB)

Video#

Small collection of 10,000 video clips of 1 minute each at 320x200 and 256 colours at 15 fps. These are to be feature indexed at 10 image frames per clip, with spoken dialogue and bibliographic record without commentary or description) is required for each. They will be indexed for structural searches. The whole collection need:

Item	Size	Comments
Bytes/video	15 MB	0.25 MB/sec assuming 4:1 compression
Raw data bytes	1,500,000 MB	10,000 articles/collection
Feature indexed overhead	200 MB	10 images/clip and 200 B/image
Bib records overhead	150 MB	500 bytes + 200%/article
Subtotal	1,500,350 MB	
Processing RAID etc.	500,000 MB	33%
Total	2,001,350 MB	= 2,000GB= 2 TB (Terabytes)

This is already compressed figure. Better compressions techniques, or a decision to lower quality, could have this figure about 1 TB.

Bandwidth#

The bandwidth needed to deliver the amount of data above becomes a serious consideration. The delivery requirements for one of the sample objects considered above are:

Object	Typical Size
One text article	10 MB
One image	300 KB
One audio clip	600 KB (1 KB/s)
One Video clip	1,500 KB (250 KB/s)

These sizes have to be compared to the available capacities of delivery channels:

Delivery channel	Typical capacity
Internal PC disk channel	33,000 KB/sec
24 X CD-ROM	3,600 KB/sec
Ethernet LAN (10 Mbps)	1,000 KB/sec
ISDN connection	128 KB/sec
56.6 modem	5.7 KB/sec
28.8 modem	2.9 KB/sec

(Source : The digital library tool kit, 2nd ed. , March,2000 Sun Microsystems. Pp 58-63.)

2.3.1.4 Processing Devices

The amount of processing for delivery is not that large for text and images. But the requirement to move video or audio data from disk storage to a network connection requires significant processing power, Special hardware, such as video server is required for this process.

2.3.1.5 Other Hardware requirements

Some other hardware required for digital library are Routers, Modems/ISDN/terminators, Tape/Disk back-ups, UPS, Printers, Fax machines, Telephone system etc.

2.3.2 Software Requirements

At each client and server, there are operating system and the application software to be considered.

For the clients, the operating system is likely to be either a version of Microsoft Windows, system based on Java virtual machine of Linux and the application software is likely to be supplied by the supplier as the whole automation package.

For the server operating system, there are alternatives like Microsoft Windows NT, Linux or UNIX. The major decision to be made here is choosing the basic application software and all its components. The package may be the system supplied by a single vendor or it may be a system with components added on to an open architecture framework.

The types of software which are likely to be needed for digital library are basically for various activities related to digitization.

- a. Integrated Library System : The systems that automate all the major functions of the library are available in the market. Such software takes care of the following activities also.
- b. Capture : This software allows capturing either audio or video from a suitable peripheral device in a computer file. Image capturing from documents is usually performed by the software that comes with the scanner. For example: for Text OCR and for Images OCR/Acrobat .
- c. Manipulation : This programme allows to manipulate images, sound files and videos. Photoshop (images), Acrobat (documents), Wavestudio (audio), Director (video and animation) etc are commonly used software for this purpose.
- d. Delivery : This is generally software added to the client (eg. Browser in the form of a plug-in) to support the playing the particular data format. Like Real Player, Real Audio etc.
- e. Web Servers : Web servers and other programmes provide the essential connection between the integrated library system and the rest of the world like Apache Web Server, which is the most popular public domain server
- f. Web publishing : To upload completed web pages to a server web page authoring tools with publishing capability are required
- g. Web Access : This software can prevent browsers from accessing undesired sites.
- h. Information Retrieval : The search engines may be used or may be connected to an Integrated Library System to provide a fully searchable collection. Some of them are Alta Vista (web search engine), Verity (free text search engine).

-
- i. Database Management System : This provides the basic storage and retrieval functions. All library systems have a DBMS underneath them. Sometimes they are standard ones like Oracle, Sybase etc. and sometimes they are specialized for the particular function.
 - j. Digital Library/Multi-media Modules : This can be stand-alone systems or can be an integrated part of the Library System or DBMS. Some of them are Image server (for digital documents), Magnifi (a multi-media web search engine, Media Asset (stores multi-media objects).
 - k. Metering : This software measures traffic on a web site mostly by analyzing the log. They have report generator and some allow custom reports to be produced. Some of them are Aria, Web challenger etc.
 - l. Rights Management : This software offers ways of controlling access to content and keeping track of who uses it for what. It also provides ways of protecting the content from theft or misuse. Some of them developed by Inter Trust, Inc and DigiMarc. (Provides digital watermarking software for all types of media).
 - m. Content delivery : Electronic delivery of content is becoming more possible and commercially feasible. The publishers now have special sites which deliver the electronic content of monographs either to computers or directly to e-book hardware. E.g. Netlibrary is for on-line delivery of books

2. 4 Support of the Staff members

Implementation of digital library becomes fruitful if the positive aspects of job advancement and new skills are emphasized. The staff should be treated as a project team and staff training/re-training should be organized to enable them to handle technical and professional activities and smoothen the whole process. The staff should be technically competent as well as having customer-oriented approach. The staff should not only know about how to handle the technology but also be aware about the concepts and issues involved in digital libraries, electronic publishing and other related aspects. The roles and responsibilities of the staff change dramatically and should be redesigned to suit the changing environment. The job becomes more challenging and therefore upgradation of the skills and knowledge become essential. The librarian needs to be practical and realistic in the approach in handling the human resource at all levels of the organization. The fullest support and cooperation of the library staff, programmers and other technical staff and the authority becomes essential to be successful in handling the transformation process in the changing environment.

2. 5 Customer Orientation

The aim of any transformation or redesign is ultimately to satisfy the information needs of the valuable customers. Therefore, it is equally important to see that customers also understand what are the new services and facilities for them. Marketing strategy of the library should also be revised to get more and more users for the materials as well as more and more material for the users (Ranganathan's laws). By meeting the expectations of the customers, a library can meet the ultimate goal and that is customer satisfaction - the reason for the existence.

3. Standards

Since there are many digital library projects going on worldwide, it is very important that standards are employed so as to make interoperability and data exchange possible. Standards are required basically to unify the representation, manipulation or transmission of some piece of information so that two or

more different systems can understand the same way. The standards are the basis of interoperability, portability and modularity. Many such standards are available and -a few of them are mentioned below.

3.1 Bibliographic Standards

These are concerned with the description of the material, both as to its content and its physical and descriptive attributes. MARC, Dublin Core, BIB-1 etc. Text encoding initiative, Metadata, AACR, Classification schemes like DDC, UDC etc, Electronic Archive description- (EAD) etc. are good examples of such standards.

3.2 Standards to Record structure

These define the physical and logical structure of the record that holds the data. These formats are for exchange of data between systems and are not intended for human consumption. Such standards are ISO 2709/Z39.2, XML, HTML etc.

3.3 Encoding standards

These concern with the way individual characters are represented in the files and records. Unicode and ASCII are such standards.

3.4 Communication Standards

For the computer systems connected with each other to pass the message back and forth the protocol of the Internet is required, i.e. TCP/IP. It is for controlling the creation of transmission paths between computers on a single network and between different networks.

3.5 Protocols

These constitute the language of the message passing between systems via TCP/IP or other protocol. like HTTP, FTP, Z39.50, ISO 23950, Z.39.63 (ISO-10160/1).

3.6 Formats

They are listed below by their file name extension. A search on the Web for three letter groups will bring up many documents describing the formats, use and their restrictions.

Images : .BMP, .TIF, .GIF, .PNG, .WMF, .PICT, .PCD, .EPS, .EMF, .CGM, .JPG...

Animation : .ANI, .FLI, .FLC

Video : .AVI, .MOV, .MPG, .QT

Audio : .WAV, .MID, .SND, .AUD

Web pages : .HTM, .HTML, .DHTML, .XML, .HTMLS

Text : .DOC, .TXT, .RTF, .PDF

Programs : .COM, .EXE

4. Conclusion

Apart from the requirements discussed above for managing the digital library successfully, the ethical concerns like Fair Information Practices and Property Rights are two main moral dimensions of information system. Development of computer networks such as internet and the digitization of information are posing new challenges to copyright. Digitization has opened the door for recording and compressing all sorts of data and copyright works. Moreover, digitization allows not only reproduction in multiple copies but also manipulation of data, images etc. The debate is going on about reevaluation of copyright act.

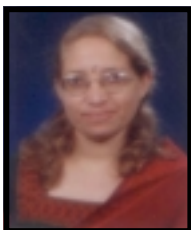
The ethical concerns of cryptography also revolves around the intellectual property and copyright issues and hence a matter of information access. In fact, cryptography seems to be the basis for implementing copyright and access authorization in digital environments. It can provide new means of protecting intellectual property in the digital world. It is advocated that digital library should develop its own policy for ethical related issues. . These policy areas can be as follows:

- Information rights and obligations: Spell out library privacy and due process policy.
- Property rights: Clarify how library will treat the property rights
- Accountability and control: Clarify who is responsible and accountable for information acquisition and delivery.
- System quality: Identify methodologies and quality standards to be achieved.

5. References

1. Sun Micro Systems. White paper, 2nd ed.
2. Arora, Jagdish. Pre-conventional tutorial on developing digital collections, INFLIBNET's International CALIBER, 2003 held at NERF, Ahmedabad.
3. Arifa, K Ethical issues in cryptography and information security: concerns for digital library, In Proceedings of INFLIBNET's International CALIBER, 2003 held at NERF, Ahmedabad

About Authors



Ms. Shweta Pandya is the Librarian of GLS Institute of Business Management and Computer Technology, Ahmedabad, India. She holds MLISc with 1st rank, CIC, DIM and qualified UGC-NET. She is the recipient of two best paper presentation awards and earlier worked with British Library Ahmedabad.

E-mail: shwetapandya@yahoo.com



Ms. Shreyasi K. Parikh is Head Librarian of National Institute of Fashion Technology, Gandhinagar, India and holds M.A (Economics) M.Lib.Sc, CIC and PGDCA. Earlier she worked with Indian Institute of Management, Ahmedabad.

E-mail: shreya_parikh@hotmail.com, shreyasiparikh@niftindia.com