

---

## NEEDS FOR DIGITIZATION OF LIBRARY : TECHNOLOGY & SUPPORTING FORMAT

Dipak Krushnarao Bhalekar

Prashant P. Deshmukh

### Abstract

*Digital conversion of Library materials has advanced rapidly in past few years. Digitization is one of the hot topics in librarianship today. To build a "Digital Library" requires that the content of a collection be available electronically. The rhetoric of the information highway has provided the impetus to convert many existing paper-based (or sound and video) collections into new digital media. Digitization of all categories of materials is now possible with the advances in digital technology. The assumption is that digital collections will be more accessible to a broader range of users, presumable through networking technologies, and that there are new efficiencies to be gained in resource-sharing and for preservation. The topic discuss the concept of Digital library, Component, Technology, Functions and Supporting Format of Digital Library.*

**Keywords :** Digital Library, Information Technology, Standards and Formats.

### 1. INTRODUCTION

The **Dr A P J Abdul Kalam**, President of India, International Conference on Digital Libraries: Knowledge creation, preservation, access, and management on 24-27 February 2004 keynote address conclusion, said that Digital Library is a national mission. We should see that all the schools, colleges, and universities digitize their libraries in their own native languages and connect to the outside world within 4 years. We have to ensure availability of fibre optic cables, satellite communication and wireless infrastructure especially in remote areas. It is also essential to realize high bandwidth technology like Multiple 10 Giga Bits connectivity across the country.

The topic of " Digital Libraries" is currently the subject of world wide debate and research. To some simply suggests " Computerization of traditional libraries" to others, who have studied library science it calls for carrying out the function of libraries in a new way encompassing new types of information resources, new approaches to acquisition, new methods of storage and preservation, new approaches to classification and cataloguing, new modes of interaction with and for patrons more reliance on electronic system and networks and dramatic shifts in intellectual organizational and economic practices.

The handling and use of rare books, fragile books and non-book materials are not new problems to libraries. In recent years libraries are acquiring non-book materials and have been developing special collections at an accelerated rate. As a result of this many university libraries have established a separate unit of special collection. This collection covers rare-books, maps, manuscripts, pamphlets, posters, broadsides, newspapers, musics, non-musical recordings, micro facsimile, microfilms, university archives, slides and films, book plates, pictures, painting, graphic materials, coins and currency, stamps, little magazines, clipping files. Etc.

### 2. NEEDS FOR DIGITIZATION

Most of India's ancient manuscripts were written on bhurj patras. Ancient India was well developed in the subjects like Astronomy, Mathematics, Space Science, Medicine (Ayurveda), Defiance, Political System,

Economics etc. Unfortunately many of these valuable manuscripts were lost due to improper care. Some of them were destroyed due to water, fungus, fire and some were purposely destroyed by foreign invaders. Book is one of the medium which deal with comprehensive knowledge and information. Preserving the books and other reading materials is a burning question which the librarians of past and present are facing. Therefore, it is urgently needed to find out some permanent solutions to preserve the library materials.

### 3. OBJECTIVES

The main objectives of the digital library are :

- To capture, store, manipulate and distribute information,
- To introduce and provide new services,
- To provide need-based and retrospective search services to the user,
- To have large number of database in CDs'
- To avoid routine and redundant activities ,
- To provide facility for networking and resource sharing,
- To access national and international journals which are being published only in machine-readable form,
- To digitize the documents for preservation and for space saving,
- To support library functions such as circulation, serial control, acquisition control, stock maintenance and other routine office works and developing in house database, and
- To improve the cost effectiveness of library operations.

### 4. COMPONENTS

Digital libraries require well-tested and proven information technologies including the multimedia kit. Much of the work in digital libraries is achieved through E-mail service, by participating in Usenet(s), by accessing the database or servers through network, little Internet. Locally developed databases will contribute a lot to develop digital libraries.

In other words, the components of digital libraries are :

- 4.1. Data-Technical Data : This consist of books and journals stored in digital form in a computer disk store. These are two ways to storing this information. One way is to photograph a page and Scan the image with a scanner. This form of storage is called a bit mapped form. It is practical way of storing old manuscripts, texts and journals. The practical way of storing old manuscripts, texts and journals. The image of the page may be retrieved and displayed on the video screen of the computer. The other way of storing a text is to present each character by its ASCII Code.
- 4.2. Numeric Data : Consist of tables of various types such as physical property data of various materials from experiments, astronomical table's etc. Such numeric data stored digitally may be used by curve setting programmer spreadsheet programmes, etc.
- 4.3. Graphic Data : May be photographs, maps, drawings etc. The simple way of storing such data is to scan the image and store it as a bit patterns. Data should in this form eases retrieval.

- 4.4. Photographs : Colour and Monochrome are stored in bit mapped form using compression algorithms to reduce storage space.
- 4.5. Audio-Data : Audio-data is digitized compressed using a commonly accepted standard comprehension algorithms & stored. A musical scores may also be coded and stored with Audio-data.
- 4.6. Video Data : Audio-data is digitized compressed using a commonly accepted standard comprehension algorithms and stored. A musical scores may also be coded and stored with Audio-data.
- 4.7. Indexing: Indexing and inter-linking multimedia data is extremely import and for ease of retrieval. Keywords in textual documents are selected and linked to related words with logical links by appropriate software. This is called hypertext. For materials on other media also related elements are selected and linked in what is known as hypermedia.
- 4.8. Linking : The information collection of the digital library will normally not be stored in one computer. It will be distributed in may computers known as servers. All these servers will be linked by high-speed communication links. The fact that the information in the digital library is distributed need not to be known to a user as it is not relevant from his/her point of view. A user gets easy access to the information based on his/her request regardless of its geographical location.
- 4.9. User : Access the library from anywhere using a terminal or a computer connected to the network to which the information servers of the library are connected.

## 5. TECHNOLOGY

Until recently, not many people were happy to read from a computer screen. Quality of display was one of the factors behind it. Recent advances in the field of Information technology have improved not only the quality of computer display but also the font displayed on them. Besides, a variety of software are now available which are capable of manipulating and rendering information. Personal computers are getting cheaper and cheaper not only in the developed countries, out are also available at an affordable cost in the developing countries. With the availability of computers, capable of computing at a very high speed and having large disc storage space, it is feasible to digitize and store information in the form of high quality graphics, colour images, voice signals and video clips at a low cost.

## 6. HARDWARE OF DIGITAZATION

In the early years of computing, computers were so expensive that it was difficult to purchase by libraries. Today, computers are cheap and middle-sized libraries in India are having computers. Apart from this, in the past, lack of hardware and absence of proper technology was also a hindrance in creating a digital library. In the present time, however, we have many hardware equipments available for digitization process. Some of these are mentioned below :

1. Flatbed Scanner : Till today, the flatbed scanner is probably the most common piece of equipment available having compatibility with any personal computer. The scanner has a glass plate onto which the source document is laid face-down. The charged coupled device (CCD), which allows to convert analog to digital, moves beneath the surface of glass and records the reflected light as an array of pixels. Flatbed scanners are both quick and economical to use. In the past their speed is increased and only single pass is required under the document. These flatbed scanners are often bundled with their own scanning software and sometime with Optical Character Recognition (OCR) software. The cost of these flatbed scanners have come down dramatically and now available at the rate as low as five thousand rupees. However, the main disadvantage of the

---

flatbed is that it relies on contact scanning, and is thus unsuitable for rare or fragile material. It also requires the document to be placed flat on the glass, so any curvature of the source material will distort the image.

2. Sheet feeders – Sheet feeders are very useful in case a library is having loose leaf documents for processing. The benefit of this type of scanner is its remarkable speed and one can leave it churning through the documents unattended. However, one can not load more than 100 pages into the feeder at one time. Yet one more demerit of the sheet feeder is that it is unable to deal with fragile or brittle items. Furthermore, these are not suitable for oversized documents, though various models of these scanner are available in the market.
3. Drum Scanners – These are called as drum-scanner because the source document is actually attached to a glass drum. Though these scanners offer the highest resolutions but rarely found outside a dedicated scanning agency because of their high cost.

Several other kind of scanners are also available such as slide scanners and Microfilm scanners. Some of these, as mentioned above, are extremely cheap. Quality of scanning is normally in resolution, measured in dots per inch. For the textual material 300 dpi is a minimum quality of scanning, and 600 dpi is considered a good-quality resolution.

4. Digital Camera – The digital camera is particularly useful in case of fragile or brittle items as these items can not bear the pressure of flatbed scanners. Many vendors have launched their products in the market over the past few years, for example, Kodak, DC215, Kontran Progress 3012, Minolta, etc. Some of these are in use in libraries in India and overseas. For instance, the central library of University of Hyderabad (UOH) is using a model of Minolta for digitization project of theses and dissertations. Large projects such as those running at the National Library of Scotland, the Arnarnaganean Institute in Iceland, British Library, etc also uses high level cameras which are capable of giving an effective resolutions. However, there are some inherent problems with the digital camera in general. For example, the scan time can be quite long meaning there by lighting conditions have to be maintained throughout potentially damaging procedure particularly when using warm lighting on documents. Also the are cumbersome pieces of equipments, which most of the time require some form of stand or cradle. Furthermore, these are not effective in the digitization of books that could not be opened beyond 45 degrees because of delicate binding.

## 7. DIGITIZATION SOFTWARE

Software for digital imaging covers three main area.

- Capturing the image
- Processing the image
- Delivering the image

Initially, we will need a piece of software having interfaced with the peripheral device used for scanning such as the camera or scanner. Important point to take into consideration when choosing the capturing software is the type of output file formats. This can be extremely significant particularly when it comes to create master image. TIFF (Tagged Image File Format) is perhaps the most important image format available at the moment. It is widely used as the cross-platform and archiving format in most digitization projects, particularly as it allows for high-quality images to be saved without any loss in the original capture. Moreover, it is not tied to any particular scanner or display. Conversion from TIFF to other formats is relatively straight forward. Furthermore, images in TIFF are the best suited for universal

---

accessibility and high resolution archival storage. However, the software should also be capable to produce JPEG (Joint Photographic Experts Group) derivatives for display on the World Wide Web. The JPEG and the GIF (Graphical Interchange Format) are the most popular image files delivered via the web and are ideally suited for displaying in all browsers. Many digitization projects create JPEG or GIF files format from their TIFF masters for subsequent delivery via the web. This is what know as processing the image and making it suitable for the end-purpose.

Once the actual conversion process over, the most important issue is the eventual delivery of the digital collection. In this step, primary emphasis should be given on what does the user want to see? The catalogues should be searchable and browsable. Facility for simple and complex searches should made available. Navigation aids should be both intuitive and apparent, and consistency maintained throughout. Additionally, there should be online help. On the one hand, the interface should be aesthetically pleasing while on the other hand, it should be open to accommodate the latest developments in technology.

## **8. FUNCTIONS**

The primary function of digital libraries is to enable searching of electronic collections distributed across networks, rather than merely electronic repositories from digitized physical materials. A digital library should have the following five integrated functions :

1. Creation and capture of content process of turning : Integrating and importing multimedia objects into digital information – digital library content from many diverse sources different type to format.
2. Storage and Management : Process of organizing and managing digital content to ensure that it is available to users whenever they want access to it. Providing stable, reliable storage and management of data in an open environment across a broad spectrum of platform.
3. Search and Access : To be able what the user need quickly.
4. Distribution : To enable information owners to distribute their material in numerous ways : through existing networks, clients/server business solutions or the Internet.
5. Right Management : Providing extensive access control capabilities enabling content owners to capitalize on their protected assets.

## **9. DIGITIZATION FILE EXTENSION AND USE**

You see a file somewhere on the 'net with a name like "library.xyz" and you would like to download it and use it on your computer. Will it work? What does the file extension xyz mean, anyway? This paper provide answers to these questions.

It is not meant to be read straight through; rather, it is meant to be a reference. An extension may be looked up either with the "find in page" option of your browser or by appending a "#xyz" (without the quotes and with xyz replaced by the extension in question) to the "go to" or "URL" field on your browser. There is also a Search Interface that will return not only the specific extension sought but also other entries that reference it. Be aware though that it assumes familiarity with the computer basics discussed in the topic.

You may notice that most extensions are three letters (or fewer) long. This is due to a historical limitation of the operating system called CP/M (that was later inherited by MS-DOC). In fact, the whole concept of file extensions comes from CP/M. Most modern operating systems do not attribute any special meaning to the "." (period, or dot) character.

Be aware though that there is no standardization to filename extension usage, and many different people have used extensions to apply to many different things.

### Supported Formats

File extensions and Full form	Extensions
The <b>audio interchange file format</b>	.aif , .aiff, .aifc
The <b>Adobe Font Metric</b> file is used to provide information on Adobe font	.afm
<b>Aldus PageMaker</b> document	.apd
<b>video</b> file will typically store a movie clip	.avi
Microsoft <b>bitmap</b> image format. This is a fairly simple bitmap format that can be viewed on many different platforms but is directly supported by very few browsers. It is not used too much on the web (fortunately, as it is very inefficient) with GIFs, JPEGs, PNGs, and even X-bitmaps being preferred as they are much more portable. It is somewhat similar to the pict format.	.bmp
A source file written in the C++ programming language	C, cpp, & cxx
A <b>command</b> file is a program for a CP/M, MS-DOS, MS-Windows, or possibly even an x86 GEOS machine. It will probably not work on a platform other than the one for which it was designed (including the different variants of MS-Windows). Similar to, but usually smaller than, an EXE File.	.com
This is a rather generic sort of extension indicating some sort of document, usually in simple ASCII; of particular note though is the fact that both Frame Maker and MS-Word often save their native format. files with this extension, and such a file may only be read with FrameMaker or FrameViewer (if a Frame document) or MS-Word (if an MS-Word document). Neither of these programs are available for all platforms, and even when they are available are usually not free. To complicate matters even further, there are several different versions of MS-Word in common use that cannot read each others' files. Thus in general this format should not be viewed as portable, and chances of reading it on any particular platform are slim. Documents saved from either FrameMaker or MS-Word that are meant for other machines should probably be saved in MIF or RTF formats (respectively) in any case.	.doc
Adobe eBook download file	Etd
<b>executable</b> is a program for a CP/M, MS-DOS, MS-Windows, or possibly even an x86 GEOS machine. It will probably not work on a platform other than the one for which it was designed (including the different variants of MS-Windows). Similar to, but usually larger than, a com file.	Exe
<b>general image format</b> file is a representation of an image or simple animation	Gif
A <b>hypertext markup language</b> file contains hypertext capable of being read and interpreted by a browser. The bulk of the world wide web is in HTML format.	html & htm
A <b>Java archive</b> should be usable on any machine with Java	Jar
image format optimized for "natural" images developed by the <b>Joint Photographic Experts Group</b>	jpeg & jpg

This is a generic sort of extension indicating "library" data; of particular note though is that native Amiga font files often use this extension.	lib
A <b>musical instrument digital interface</b> file is an audio file that will possibly work with any device that is MIDI aware.	midi & mid
A <b>QuickTime movie</b> can actually be used to store both movies and sounds	mov & qt
The <b>Moving Pictures Expert Group</b> devised formats for storing both movies and sounds. A file with this extension could thus be either an audio file or a movie file. Both types are supported by many different platforms, with perhaps the audio type being a little more recognized. Both types also utilize lossy compression that is designed to take advantage of the limitations of human vision and hearing. Regardless the compression still occasionally does weird things producing "MPEG compression funnies", particularly during periods of rapid movement in video. MPEGs are not limited to computers; some digital satellite television transmissions use an MPEG format, and DVDs use an MPEG format. MPEG has nothing to do with either JPEG or JBIG in spite of the similarity of names. MPEG does have some relation to QuickTime, though, and the next version of each will work to merge the two. If there is a number at the end of the extension, it refers to the specific generation of the MPEG format in use within the file.	mpeg, mpg, mpeg2, mpeg3, mpeg4, mp2, mp3, mp4, m3u, & m4u
The <b>Portable Document Format</b> is a format for storing all manner of documents. PDF documents can contain sounds and animations in addition to simple images and are binary in nature. PDFs are also fully capable of containing hypertext, but unlike HTML will guarantee proper printout.	Pdf
An Adobe <b>PageMaker</b> document	pm, pm3, pm4, & pm5
A Microsoft <b>PowerPoint</b> document	Ppt
The <b>PhotoShop Data</b> image format was developed by Adobe	Psd
The <b>tagged image file</b> format is a loss less format for storing image data. TIFF is a popular output format for scanners and other similar devices.	tiff & tif
This is a generic sort of extension indicating a simple text file (usually ASCII). It can be readily used on virtually any computer, although sometimes character set differences will require slight conversions.	txt & text
A <b>WordPerfect</b> document	wp, wp4, wp5, wp6, & wpp
An Excel spreadsheet document; can be read on any machine with Excel	xls
A file that has been compressed with either the zip or pkzip program will get the "zip" extension.	zip

## 10. CONCLUSION

It is clear that emergence of Digital Libraries will severally affect the traditional function of libraries. The affordable prices and wide availability of quality & canning equipment has made a reality to convert original documents into digital images. Digitization also helps organize collections that have become unmanageable because of Size, Physical location etc to ensure that it is available to users whenever they want access to it.

---

## 11. REFERENCES

1. Bhalekar, Dipak Krushnarao and Deshmukh, Prashant (2005) Digital Library: Technology & Services, Kochin: Caliber, Feb 2005.
2. Dr. A.P.J. Kalam (2004) International Conference on Digital Libraries (ICDL) , Mumbai: SISCON V.24
3. Gupta Rajiv (2003). Preservation of Digital Information : Some Aspects. *IASLIC Bulletin*. V.48(4). P. 207-208
4. Deshpande, N.J. AND Panage, B.M. (1999) Digitization of Library Materials. University News V37(38) P.9
5. <http://www.Saugus.net/computer/search>.

### About Authors



**Sh. Dipak Krushnarao Bhalekaris** working as an Assistant Librarian in B.N.College of Engg., Pusad Ta. Pusad, Dist. Yavatmal, Pin-445215 (MS) India.

**E-mail :** [dkbhalekar\\_bnce@rediffmail.com](mailto:dkbhalekar_bnce@rediffmail.com)

**Dr. Prashant P. Deshmukh** is currently working as a Librarian in P.N.College, Pusad Ta. Pusad, Dist. Yavatmal, Maharashtra, India

**E-mail :** [ppdeshmukh\\_bnce@rediffmail.com](mailto:ppdeshmukh_bnce@rediffmail.com)