DIGITIZATION INITIATIVES AND UNIVERSITY LIBRARIES IN INDIA

S N Singh

Abstract

This paper discusses the various aspects of the conversion of in-house library materials into electronic format, which leads toward the creation of digital library. Also describes the digitization initiatives that have taken place in India. Objectives of digitization have been correlated with the mission of university libraries. Deals with various image capturing and processing technologies and suitability for a university library. It tries to identify the areas to be covered under digital conversion program of in its first phase.

1.   Introduction

Information Technology is moving so fast that is making difficult to cope with changes & challenges. Information activities have undergone rapid transformation from conventional methods, consequent upon introduction of new technologies. To meet the tremendous information explosion and high demand of information, libraries are now taking the advantage of digital technology available at the cheaper rate. Now every library and informational professional in India is talking about digitization and digital library. 15 years back LIS professional only spoke of automating the library activities. Recent advances in computer' storage & processing, communication technologies, e-products, networking, internet use have brought a revolutionary changes in functioning of the libraries.

Digitization started with the introduction of computers in library activities in 1980s in India but it was limited to creation of bibliographic database and house keeping operations. This conversion of bibliographic catalogue record into digital format was the initial stage, but now it has been broadened by implementing it to conversion of full text, graphics, sound, video records etc.

A lot of confusion is still prevailing among many library and information professionals about different types of libraries emerging out of the development and convergence of computer and communication techniques. So it is appropriate here to say something about these types of libraries. Traditional library is a library where the access points such as, library catalogues as well as the library collections are print-based and management is by and large manual. In an automated library, access points and housekeeping operations are computerized. Electronic library and digital library mean one and the same thing. The British people call it electronic library and the U.S. people call it a digital library. In this type of library both the access points as well as the graphic records are in electronic/digital form, whereas in an automated library the graphic records are still print on paper publications. When the electronic libraries/ digital libraries are connected via various networks particularly the

Internet, this is called a virtual library. In such a library traditional constraints of time and space stand collapsed

2. Major Digitization Initiatives in India

Some of important digitization initiatives undertaken in India are given below:

- The National Science Foundation, USA has initiated the "million book project' at Carnegie University in USA in collaboration with The Indian Institute of Science, Banglore as his Indian partner. The project is an international cooperative attempt to digitize a million seminal books in fields that happen to be in the public domain or are copyrighted but out of print, and make them accessible through the web globally. By 2005 almost one million books have been digitized in India alone.

- The national Library of India has undertaken digitization of often used books and documents such as East India Co. records, diaries etc. as well as MSS on paper and palm leaves.

- The Parliament Library has digitized all debates, questions, committee reports, biodata of present and past members of Parliament including photographs and addresses etc.

- National Mission for Manuscripts launched by Government of India in 2003 is milestone in conservation and preservation of manuscripts for posterity as well as for giving access to a scholar from any point of world through internet. India being the largest repository of manuscripts in the world has innumerable ancient MSS and rare books, which urgently need digitization. The National Informatics Centre, Government of India has prepared detailed guidelines for digitization of such materials.

- The Indira Gandhi National Centre for Arts (IGNCA) is pioneering digital treasure of Arts, digital images etc. IGNCA has been recognized as the nodal agency for National Mission for Manuscripts of Government of India.

- Indian Academy of Sciences has digitized all its journals.

- Some of Institutions and Universities, namely IIT, Delhi and Mumbai, University of Hyderabad, Physical Research Laboratory, Ahmedabad, Banaras Hindu University, Varanasi have started the digitization of their Manuscripts/Thesis /Dissertations under different projects.

- IIT, Kanpur, Allahabad University, Central Institute of Higher Tibetan Studies, Varanasi, etc are digitizing the collection of books (free from copyright) under the project initiated by The National Science Foundation, USA in collaboration with The Indian Institute of Science, Banglore.

- The National Chemical Laboratory has the digitized national collection of industrial micro organism.

3. Mission of University Libraries and Digitization

University library system exists to serve the information needs of its academic community and supports the teaching and research mission of the university. University libraries have following mission:

- Acquiring, storing, processing and distributing the information
- Support the teaching & research program
- Serve the information needs of students, teachers and staffs

- Conservation and preservation of knowledge
- Dissemination of publication of research results
- To work for maximum satisfaction of users.

Goals of Digitization more or less match with the university library mission. The basic objective of digitization is to overcome the twin tyrannies of time and space. By digitizing we go toward maximum user satisfaction. No doubt it will support the learning, teaching and research of the university. In the light of mission of university library we find similarity to the owing points:

- Preserve to use for posterity
- To Capture, store, process and disseminate information in digital form
- Save the space and time both
- Facilitate for networking and resource sharing
- Make available to remote users
- Maximizing user' satisfaction by providing multiple and easy access
- Foundation of new value added library service

We can overcome with the problem of space in accommodating the materials which every university library is facing presently.

4. Digitization

Digitization is the conversion of materials or analog media into electronic form for creating digital collection. Analog media includes books, journal articles, micro form, photos, audio, video, etc. Simply it is high speed data transmission technique. In digitization, work is compressed in digit (0,1) form. 0 and 1 is a symbolic form about polarization of ions in magnetic media, where we store the data. In the case of digitization of text matter the images should be in bi-tonal i.e. black and white.

Digitization is necessitated for both accessibility and preservation. By creating and building the digital Information we will be leading towards digital library. Digital collection is a most important component of the digital library. A digital library is not a single entity although it may have digital contents created in-house or acquired from outside in digital format and loaded on network.

4.1 Hardware and Software Requirement

Today we have relatively cheap hardware equipments required for digitization, which are affordable by University libraries. Hardware and Software Requirement for digitization are given below:

Computer system	Pentium IV Processor with 1 GB RAM and 40 GB Hard disk, and Windows 2000/XP or Unix Operating System
Scanner	Flatbed Sheet Feeders or Plannetary
Digital camera	Nikan D100/ Kodak/ Minolta/Sony/ Canon etc with minimum 5 Mega Pixel
Image processing/editing	Adobe Photoshop/ Kodak Imaging software
File Compression software	MrSid Image Compression software
OCR Software	Prime Recognition

Descriptions of the some of the common and important hardware equipments available for scanning are given below:

4.1.1 Flatbed Scanner

Presently this is most popular and common scanner, which have the compatibility with any of the personal computer. These are both quick and economical to use. This is available at quite reasonable rate in the market. Its cost has dramatically come down to Rs 5000/- . The scanner has a glass plate onto which the source document is laid face-down. The Charged Coupled Device (CCD) allows converting analog to digital moves beneath the surface of glass and records the reflected light as an array of pixels. The Flatbed scanners are often bundled with their own scanning software and sometime with OCR software.

4.1.2 Sheet Feeders

This is very useful particularly for the scanning of loose leaves. It has a remarkable speed and one can leave it churning through the documents unattended. 100 pages can be loaded at a time. This is not suitable for oversized documents.

4.1.3 Drum Scanners

These are known as drum scanner because the source material is actually attached to a glass drum. These offer the highest resolution but not affordable by the libraries because of its high cost.

4.1.3 Digital Camera

Digital camera is very useful for scanning of fragile materials like manuscript, rare book., as these materials can not bear the any touchable pressure of scanning. Some of the important available in the marker are: Kodak, DC 215, Minolta, Kontran Progress 3012 etc. some of these are being used by the libraries for digitization of their manuscripts, theses, dissertations. Lighting conditions has to be maintained throughout the process. Although it takes much time in comparison to other scanners.

Selection of imaging equipment for digitization is very important, while selecting cost plays much important role. University libraries may use digital camera (ie Nikan DC100, Sony, Minolta) with minimum 5 mega pixel for their manuscript and rare book collection. For other documents flatbed scanner is useful as it is affordable by the libraries. It has a good resolution also.

5. Consideration & Preparation Before Starting RECON

It is not possible for any University library particularly developing country like India, to digitize (full text) the whole collection in a single phase. Huge fund and time will be required for such project. Conversion of such existing materials/records into electronic/digital format may be recogniged as 'Retrospective Conversion' in short RECON. We should adopt some strict measures before we proceed for conversion. The time and cost required to create digital records for the remainder must be taken into consideration.

6.   What is to be digitized?

University library has to decide what is to be digitized and what can be preserved on other media. Whether there is a need to include whole collection for digitization or a portion of it. The decision may be taken dictated by the amount of fund available. A traditional university library collection generally consists of the following:

- Text books/ reference books
- Manuscripts and rare books
- Serial publications
- Theses and dissertations
- Technical Reports and Annual reports.
- Newspaper and Magazines
- Audio/video cassettes
- Microfilm/microfiche

By building the digital collection University library may proceed to develop the digital library. Which may include electronic journals, electronic books, full-text article databases, CD-ROM databases etc?

There are many problems: managerial, technical and financial, which are to be considered & solved at adequate time by concerned authority. Conversion costs money to create, process, update and store the record.

- Deciding the areas covered under digital conversion
- Deciding the priority among areas of the collection to be converted
- Time frame, determining the speed
- Cost factor

After careful study of need of users, we should initiate for active collection and the collection free from copyright issue. Following items of university collection may be taken for digitization in order given below:

- Manuscripts and Rare materials
- Thesis / dissertation
- Technical reports and annual reports
- Special collection, if any

Since Manuscripts form a significant resource material of the university library collection for study and research, it should be at our top priority. Digital preservation of old rare documents and manuscripts is essential because it contains valuable information about society and culture. By digitizing them we should try to preserve and make available to all through network. Thesis and dissertations are another important component of any university library collection which needs to be digitized at the priority basis. As these may not be available to other library users because of unavailability in print or other media. Digitization of these will be proved fruitful for users. Technical reports and annual

reports issued and published by institutions are grey literature, containing valuable information not available through any other channel. University may shift their priority of digitization based on the needs of clients and nature of the institutions. All these does not relate to copyright issue so digitization of such materials is easy.

7.   Identify RECON option

There are several options available to libraries for retrospective conversion. Each one has its merits and demerits. We should analyze and choose the option carefully. The options to carry out RECON are as under:

- In-house conversion by existing staff
- In-house conversion by outside agency/contract workers
- Outsource off-site conversion
- In-house conversion using existing staffs and contracts, a combination of above two

Cost is major factor, which is to be considered in choosing the option. For each option chosen for analysis it is necessary to determine the costs for the following:

- Funds available
- Staff requirement
- Equipment
- Vendor charges
- Location & time

Among the alternatives available, in-house digitization using existing staffs and contract workers (trained) alternative is more practical. The Library proceeding for retrospective conversion would require additional hands to carry out the job. One major benefit with this option is that documents need not to send outside. There is always risk of missing of materials and user's loss in case documents sent outside for conversion. Among other factors, it is in our command and supervision, as per our need. Without any extra effort work of physical verification is done, simultaneously, when we input data with the help of documents in hand. One major advantage of this exercise is that a sort of stock verification is carried out concurrently without any extra effort.
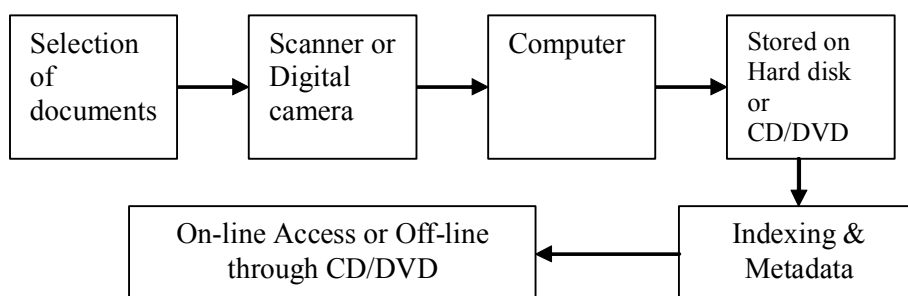
8.   Digitization Process

The existing resources of the library can be converted into digital media by scanning the materials. Flatbed scanner or digital camera may be used for image capturing or scanning. Scanning of images may take place at 300 dbi bitonal. 600 dbi bitonal is used for higher resolution. Scanning is the most popular method of converting the materials. Digital imaging technology has a number of options that can be adopted depending on the objectives, its end users, availability of fund etc.

After choosing one of the data conversion methods, as suited to particular library, we should develop plan of action. Digitization process starts with project planning with documentation followed by a linear progression of tasks : select/prepare, convert/scan, process, store, distribute and maintain over time. Digital conversion process includes:

Document

- Data/image capture
- Data Processing
- Storage
- Indexing and processing
- Retrieval and display

```
┌─────────────┐    ┌─────────────┐    ┌─────────────┐    ┌─────────────┐
│ Selection   │───▶│ Scanner or  │───▶│ Computer    │───▶│ Stored on   │
│ of          │    │ Digital     │    │             │    │ Hard disk   │
│ documents   │    │ camera      │    │             │    │ or          │
│             │    │             │    │             │    │ CD/DVD      │
└─────────────┘    └─────────────┘    └─────────────┘    └─────────────┘
                                                                │
                                                                ▼
      ┌──────────────────────────────┐         ┌─────────────────┐
      │ On-line Access or Off-line   │◀────────│ Indexing &      │
      │ through CD/DVD               │         │ Metadata        │
      └──────────────────────────────┘         └─────────────────┘
```

Selection and identification of materials/documents according to priority decided earlier is first step. A list of candidate documents for entire project may be prepared as a database. It should have identifying information, priority based on content and its copyright status. After selection each page may be scanned as a separate image file. A spreadsheet file is created for each document. The fields in the spreadsheet are used to automate certain aspects of data creation. This spreadsheet includes the item identification number, title, number of pages and notes of each document. All the resulting files are collected into a single directory. The text in the convertible document may require conversion of diacritics, or special characters, images may need enhancement, amplification or compression. Records are verified and processed/edited, if required. The information converted into digital form will need digital storage medium, which may be hard disk, magnetic tape, optical CD-Rom, or networks with workstation to access. Use of Dublin core standard is advisable to create metadata of digital resources.

9.   Output Format

The images may be obtained in three formats, namely TIFF, PDF, JPEG. All the image files of the individual pages are obtained in uncompressed TIFF and JPEG with the objective of archival. Printed Document Formats (PDF) containing the individual pages in the form of electronic book may be used for viewing and access.

10. Conclusion

Digitization projects have been important for university libraries aiming the digitization of manuscripts, theses, dissertations, special collections, which are special in nature. Some of the Research and Special libraries have already undertaken digitization projects in their full swing. Few digitization initiatives by university libraries could be seen. Yet efforts are to be made in this direction by university libraries in India. It will provide solution to their several problems, like: space, preservation of age old and fragile materials, users' satisfaction, network based full text browsing, time, etc.

Preservation issues, copyright, intellectual copy right issues, bibliographic integrity, identification, specification and standards are important issues to be kept into mind while proceeding for digitization. Substantial time must be given for performing quality control checks and every stage of processing. For more manageable quality control, materials should be processed in batches. University libraries need to build digital information sources and services using the modern technology to satisfy the users demand and laws of Library Science.

References

1.   Ansari, Mehtab Alam. Digital libraries: needs, technology & benefits. ILA Bulletin, Vol.39 no1, 2003.

2.   Dasgupta Kalpana. Digitization, sustainability, and access in Indian context. Paper presented in 71th IFLA Conference. 14-18 August, 2005, Oslo.

3.   Gakhar, AP. Digital preservation of manuscripts. In souvenir of National Seminar on Digital Preservation of Manuscripts & Rare Materials. Org. by Banaras Hindu University, Varanasi. 2005.

4.   Galloway, Edward & Brenner, Aaronl. Digital library projects at of University of Pittsburg. Encyclopedia of Library and Information Science, USA, 2003.

5.   Shaw, EJ. Building a digital library: a technology manager's point of view. J. Academic Libraries, 2000.

6.   Singh, Jagtar & Kaur, T. Virtual library: a useful mechanism for globalization of information. Progress of Library and Information Science, BHU, Varanasi. 2002

7.   Digital preservation. http://www.bl.uk/about/collectioncare/.html

8.   http://www.pib.nic.in/feature/feyr2002/f240720021.html