

DIGITAL TRANSFORMATION OF LIBRARY COLLECTION: PROCESS AND CHALLENGES IN INDIAN CONTEXT

Sumeet Kumar Handa

Dy. Librarian

THDC-Institute of Hydropower Engg. & Tech.

Tehri (Uttarakhand)-249124

Email: librarian.thdcihet@gmail.com

Dr. Sudhir Tyagi

Head, Dept. of Lib. & Info. Sci

Swami Vivekanand Subharti University, Meerut-250002 (U.P.)

Email: slib53@rediffmail.com

Abstract

This paper discusses the concept of digitization and the process of digitization of library collection. It also deliberates the challenges often faced during digitization. Digitization suggests transformation of documents and art works into digital images. Digital images here mean electronic copies of documents. Digitization is a procedure in which materials are converted from the hard copies to electronic copies. The major purposes of digitalization are: to enhance access and improve preservation of library materials. A number of challenges are encountered in the process of digitizing library materials. These challenges include human and technical problems, which have implications for planning and policy. Apart for this the paper is also put light on Indian digital scenario and tries to describe the difficulties in the digitization process. In this ICT era, digitization is an essential task to transform libraries from traditional form to contemporary, because of the current information and knowledge store, process and dissemination challenges demands to go for digital. Besides, the progress of ICT services and online services in the library requires digitization.

Key Words: Digitization, Digital Library, Transformation, Challenges, ICT.

INTRODUCTION:

The Present day information age is the practical world. The practical world is exists because there is a physical world and the former cannot exist without the latter. The physical world is exists in various formats of Information. A Digital Library is an automated or electronic Library where activities like accessioning, retrieval processes, automatic indexing and textual analysis are carried with the help of computer. The user can access the information from anywhere, anytime and information can be stored anywhere in the world. Digitization of information basically means of the process of converting paper documents such as text, graphics etc. into digital image, which can be made accessible through electronic network. We can say digitization means process

which is mean to eliminate paper work and make a transition to a paperless environment with the help of software technology. In the current scenario library operational activities are computerized because of the increasing size and multiplicity of material of the library, mainly for efficient library operation, such as administration, acquisition, circulation, serial control, cataloguing, information retrieval etc. The library which uses computers for housekeeping and information retrieval is called an “automated library”.

DIGITAL LIBRARY:

A digital library is a collection of digital documents or objects. This definition is the dominant perception of many people of today. However, Smith (2001) defined “a digital library as an organized and focused collection of digital objects, including text, images, video and audio, with the methods of access and retrieval and for the selection, creation, organization, maintenance and sharing of collection”.

Though the focus of this definition is on the document collection, it stresses the fact that the digital libraries are much more than a random gathering of digital objects. They retain the several qualities of traditional libraries such as a defined community of users, focused collections, long-term availability, and the possibility of selecting, organizing, preserving and sharing resources.

According to Wikipedia “A digital library is a collection of documents in organized electronic form, available on the Internet or on CD-ROM (compact-disk read-only memory) disks. Depending on the specific library, a user may be able to access magazine articles, books, papers, images, sound files, and videos.”

DIGITIZATION:

Digitization of information basically means of the process of converting paper documents such as text, graphics etc. into digital image, which can be made accessible through electronic network. We can say digitization means process which is mean to eliminate paper work and make a transition to a paperless environment with the help of software technology.

- i. Digitization means no new buildings are required; information sharing can be enhanced and redundancy of collections reduced.
- ii. Digitization leads to the development of Internet in digitalized based libraries. As Internet is now the preferred form of publication and dissemination.
- iii. Digital materials can be sorted, transmitted and retrieved easily and quickly.
- iv. Access to electronic information is cheaper than its print counterpart when all the files are stored in an electronic warehouse with compatible facilities and equipment.
- v. Digital texts can be linked, thus made interactive; besides, it enhances the retrieval of more information.

In the light of the following advantages, it is natural today to find more information being digitized and uploaded into the Internet or Compact-Disc Read Only Memory (CD-ROM) in order to be made correspondingly accessible globally.

Why Digitization?

There are three main needs for digitization; two or all the three of them may apply to your digital library project.

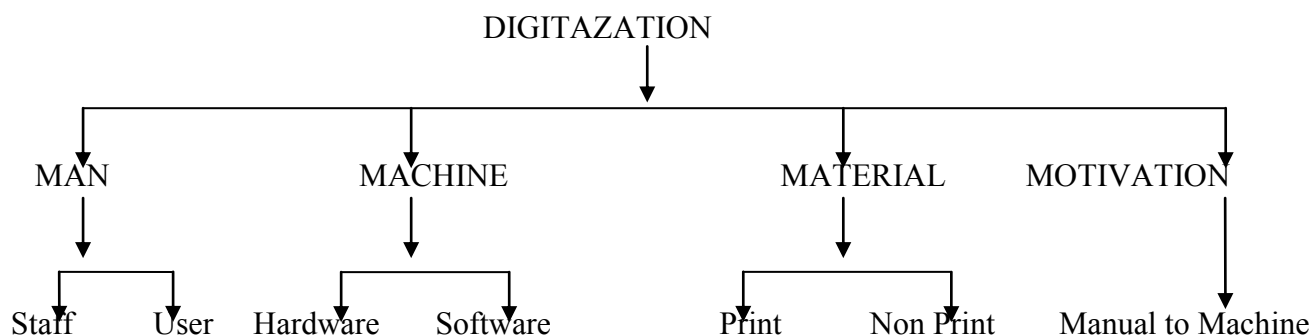
- i. To preserve the Documents: That is to allow people to read older or unique documents without damage to the originals.
- ii. To make the documents more accessible: This is to serve the existing users better; e.g. to allow the users to search the full text of the documents or to serve more users than envisaged in remote locations, example, more than one person at a time.
- iii. To reuse the documents. It means to convert documents into different formats; for example to use images in a slideshow and to adopt the content for a different purpose.

There are a number of reasons for each library to think of digitization. There are also many ways to create the digitized images, depending on the needs and uses. The need, of users for information of a high quality and varied resources, is the prime reason for digitization.

- 1- Quality preservation: Having access to good quality matter, images, sound etc., in perpetuity. The scanned images and matter are of the best possible quality and the quality remains the same whenever and, however, many number of times it is viewed. It is becoming a common practice to provide a CD-ROM with a printed book. The reproduction does not have generation loss if the original is in the digital format.
- 2- Multiple Referencing: To make the information accessible to several users at a time. Many members on different terminals can review information kept on the server in the institute or on WAN at the same time. Mirror CDs are common nowadays.
- 3- Wide Area Usage: Makes it accessible to users at a distance. The information put on the server of a large area network can be reviewed at the same time in different locations. The best example is the internet.
- 4- Archival Storage: Restoration of rare material. Keeping rare books, images or archival material in digitized format is now a common practice.
- 5- Security Measure: Valuable documents and records are scanned and kept in digital format for security. It is also considered as an enabling provision for restoration of rare material. Library of congress keeps the whole catalogue 30 miles away for safety reasons.

Digitization Requirements 4M's:

- 1- Man, (A) Staff (B) User
- 2- Machine, Hardware, Software
- 3- Material, Print, Non Print
- 4- Motivation, Manual, Machine



OUTLINE OF DIGITIZATION:

Before the process of digitization the following stages initial to discuss:

(i) **Policy enactment:** A policy is a guiding statement. The top management should enact a policy on the project. Such a policy will serve as a reference point and guide for implementing the project. The policy should contain the goals of the digitization project. Good goal setting is important for any new initiative, and digitization is no exception. The goal 'To make our materials more accessible on the web' is not specific enough. There is a need to be specific, particularly on the categories of users that will access the collection, the type of material they may be interested in, how they will use it, how many people are envisaged to use it, the planned procedure for its advertisement, and the benefit of the material to users and institutions. Contacting current and potential users is an excellent way of having clues to all these issues. One may consider sending out a survey to the project's intended audience in order to learn how they are currently using the material, and how they might use it differently if it was digitized. It may be helpful to contact other institutions that have digitized similar collections and learn from their successes and failures.

(ii) **Policy approval:** The policy should be approved by appropriate authorities before project implementation. For instance, a university library may need the approval of the university management and other funding agencies before any digitization project can be embarked upon.

(iii) **Planning, budgeting and monitoring:** This is a very essential stage. It is desirable to set up a planning committee that will draw the plan and budget for the digitization exercise. Budgets for digitization projects should include the following categories:

- (a) Salaries, wages and benefits.
- (b) Staff training.
- (c) Equipment and supplies.
- (d) Services, contracts and legal fees.
- (e) Overhead and indirect costs (including offices and workspace).

(f) Maintenance, licenses, and communications charges; and

(g) Contingency (setting aside about 10% of the total project budget for unexpected expenses).

The purposes of the digitization project, the source of fund and the amount available for the project should also be taken into consideration. At the regional or national level, effective planning for digitization can bring together all types of libraries, museum, academic/professional societies, historical societies and archives to take advantage of the exercise. In USA, the planning for digitization in the Central New York brought together all types of libraries, museum, historical societies and archives which took advantage of expertise and content. The Central New York digitization project was supported by a Library Services and Technology grant provided by the New York State Library.

(iv) Acquisition of appropriate technology: The plan drawn for the project will determine the appropriate technology to acquire. Technology here refers to all the equipment/hardware and software that are needed.

(v) Administrative decision: on the procedure to be adopted: Decision has to be made on the mode of operation, whether to just establish links with existing digital libraries or to digitize in-house or to contract it out. There is a need to establish time limit for the project.

(vi) Sensitization: psychological preparation and retraining of staff: In most places the staff will like to resist the digitization project. It is a common thing for people to resist change, just for the fear of the unknown. The library staff may fear that the success of the project may affect their jobs adversely. Those who are not computer literate may not be willing to adjust. All these categories of people have their genuine reasons to resist. It is the responsibility of the library management to educate them and allay their fears.

(vii) Copyright permission: Violation of the copyright laws should be avoided. It is not necessary to obtain copyright permission for materials published before 1922. Copyright permissions have to be obtained for materials to be digitized, particularly those that are not available in the government domain. When the copyright permission is granted, it is essential to enter the date of approval and the name of the person who granted the permission into the database.

(viii) Implementation and trial testing: At this stage it good to start with trial testing, using a few materials as samples. This will enable us to know whether the format and fields are flexible and suitable. Adjustments can be made. A pilot digitization project should start with a manageable collection. Focusing on items with consistent or standard formats (photographs of all one size or type, documents from one collection, etc) provides the best chance of success. If the trial testing is successful, the project can be commenced. Data entry is rigorous, time consuming and very expensive. Existing materials can be scanned. Modification of scanned and digitized documents is very essential; so as to minimize errors. This will enable developers to put them in appropriate formats.

(ix) Evaluation of project: The top library management needs to be making periodic evaluation of the project. This will reveal lapses that have to be addressed. Evaluation is a neglected aspect of digitization projects. Project evaluations should not just be easily quantifiable figures or an attempt to determine program's impact on the user. Several digital projects are judged by the number of items they digitize. This is really one of the least useful measures of a project's success. The number of images digitized means nothing, if they are of low quality, hard to locate in a database, or not interesting to the public. Assessing how users are using digital materials provides a more effective evaluation tool. At the bare minimum, projects should be formally evaluated based on the set goals.

PHASES FOR PROCESS OF DIGITIZING DOCUMENTS:

Cornell University Library/Research Departments (2000), provides six stages in digitizing documents for a digital library: Registering, Scanning, Optical Character Recognition, Proofreading and formatting and producing the Final Version:

i. Registering: Before scanning large number of documents, there is the need to first register them and use a filing system to keep their track. If not, you risk misplacing hardcopies, losing files, skipping steps in the process or duplicating work, perhaps without realizing it. There is also the risk of losing electronic versions of files because they have been misnamed or saved in the wrong subdirectory. Moreover, a good filing system is vital, so everyone in the digitizing team knows what he is supposed to do, and he can fill in for another person in case of absence.

ii. Scanning documents: It is necessary to clean and dust off the documents to be scanned; make sure that all the pages are present and in the right order. If the document is in poor condition, try to find a fresh copy. If it is a sheet fed scanner, cut the document open to get individual sheets to feed through the scanner. If necessary, you can rebind the documents later. If you do not want to damage the documents, you can photocopy each page and feed in the photocopy through the scanner, though this uses a lot of paper and reduces the quality of the scan.

To scan a document on a flatbed scanner, place it face down on the scanner platen or put the pages into the sheet feeder. Then, in the software, choose a setting, resolution and colour and scan each page of the document at the settings you have chosen.

iii. Optical Character Recognition (OCR): Optical Character Recognition (OCR) software converts a scanned image into a text file that a word processor can read. To do this, it must first recognize where the text is on the page. The software breaks the text blocks down into lines or into an individual character. It tries to match the image of each letter against patterns it recognizes as an "a", "b", etc. There is a problem to encounter with languages that use Latin scripts with accented characters. As a solution, you should use the OCR software that is specific for language.

iv. Proofreading: This is the act of making corrections to the document text and layout. This is done in two ways:

- a. Comparing the scanned text on the screen with the hardcopy and entering the corrections directly into the computer. The word processor's spellchecker will help in spelling errors quickly.
- b. Printing out the scanned text and comparing it with the original copy. Mark any corrections on the printout, and then enter them into the computer. This is a slower method, but may be the best option if there are no enough computers for each proofreader.

v. Reformatting: The Optical Character Recognition (OCR) software may produce a document that consists of straight text, no columns, no headers and footers. There is the need to reinsert these by hand or correct where they appear on the page. There may be also need to change the typeface, heading styles and so on, to make the document more attractive and readable. Alternatively, you may be able to adjust the settings of your OCR program to preserve the layout of the page.

vi. Final Version: For many documents, there is a need to add some information to the text so that readers can identify it easily. As for a book you must make sure that the book title, the author or the editor, the publisher and the publication date are all included. As for chapter in a book, you should include the title and the author of that chapter and the original page numbers in the printed version of the book. As for the journal articles you should include the journal title, the date, the volume and the issue number, the article title and the authors and the page numbers in the original printed journal. In other words there is the need to add Metadata to describe each document.

CHALLENGES OF DIGITIZATION:

Digitization of library collection poses a great deal of challenge to the major stake holders, that is, the library management, employees and library users. The library management has to source for fund for the digitization project. It is often easy to get sufficient fund for the project, as the required fund can be enormous. Most academic libraries face unmanageable budgetary demands. The financial constraints have to be taken into consideration. This is essential so as to be able to continue with the digitization exercise after the conversion program. Clear cut decisions have to be taken on the form of digitization to adopt. Where the fund available is grossly inadequate, the library can be linked to existing digital libraries. It is also possible for the library to select and digitize additional materials. The proportion of the additional materials to be digitized will therefore depend upon the available fund. The problem of most digitizing efforts is that of inadequate fund and not that of technology. There should also be continuous flow of fund so that the project can be functional. The importance of the project makes this a very good suggestion. The academic libraries are often located in research environments, where students, faculty and other researchers will depend on the library for timely information. Management of the digitization project entails policy initiation, setting priorities and planning. These are challenging tasks for the management. The library management needs to consult libraries that had digitized their materials so as to learn from their experiences. This will guide a lot while formulating policies on the digital project. A planning committee has to be set up. It is the responsibility of this committee to draw plans and budget for the project.

The library management will also need to prioritize the different activities involved and assign each task to a committee. Time limits should be assigned for completion of each task. The task of carrying along all the staff and guiding library users can be challenging. Some of the staff will like to resist change, particularly those that are not computer literate. It is essential for the library management to explain the essence of the project to them and arrange to retrain the employees so that they can participate in the project and remain functional in a digital library. Some of the library users will definitely find it difficult to search for materials in the digital terrain. It is good for library assistants to be available to render assistance. Orientation programs can be organized for these library users from time to time. This is very essential in academic libraries where the users can be many and are also regular users.

CONCLUSION:

Libraries are the storehouse of knowledge as they maintain the book and other knowledge resource available mostly in printed form. However, with advent of digital technology and ICT, the library scenario is changing fast globally as well in India also. Information and Communication Technology (ICT) has revolutionized the concept of libraries. Numerous services and facilities of ICT urge to Indian libraries that they should implement the ICT service and go for digitalization. Indian Government has been starting many major projects to transform traditional libraries of India towards Digital Libraries such as Digital Library of India (DLI), Nalanda Digital Library, and Digital library of Indian Institute of Management, Kozhikode and many more. Digitization has opened up new audiences and services for libraries, and it needs to be included into the strategy and policies of any institution to make best use of its effectiveness. Digitization is an intricate process with many critical dependencies between diverse stages over time. Utilizing a holistic life-cycle approach for digitization initiatives will help develop sustainable and successful project. It is hoped that the approach of the issues outlined, the software mentioned in this paper and the references to more detailed source and past project will contribute to the future success of initiating digitization of library resources.

References:

1. Witten, I. H., & David B. (2003). *How to Build a Digital Library*. Morgan Kaufmann Publishers: San Francisco: USA.
2. Raina, R.L, Gupta, D.K., and Gaur, R.C. (2005). *Library Management: Trends and Opportunities*. Excel books: Delhi: India.
3. Arms, W. Y. (2000). *Digital Libraries*. The MIT Press: Cambridge: USA
4. Nazeer, B.K. (2008). *Digital Library Architecture*. Ane Books. New Delhi: India.
5. Bhat, P.S. & Baheti, S.R. (2011). Digital Library: Issues and Challenges in proceeding of national seminar on Impact of information Technology.
6. Kumar, K. (2014). A Scientometric Study of Digital Literacy in online Library Information Science and Technology Abstracts (LISTA). *Library Philosophy and Practice* (e-journal).
7. Kumar, S. (2016). Digital Library: concepts and usages in International conference from ownership to access: leveraging the Digital Paradigm.
8. http://dir.yahoo.com/Reference/Libraries/Digital_Libraries/
9. <http://librarydotcom.webs.com>

10. <http://www.google.co.in>
11. www.utm.edu/staff/jfieser/class/120/6-knowledge.htm
12. https://en.wikipedia.org/wiki/Digital_library
13. <http://www.nap.edu>
14. <http://www.dli.ernet.in>
15. <http://etd.ncsi.iisc.ernet.in>
16. <http://www.ignca.nic.in/>
17. <http://www.nalanda.nitc.ac.in>