

Digitization of library resources and the formation of digital libraries: Special reference in green stone digital library software

Sanjay Singh

Library Assistant, Central Library, Banaras Hindu University, Varanasi, Uttar Pradesh, India

***Corresponding Author:**

Email: sanjaybhu196@gmail.com

Abstract

This paper discusses the new activities, methods and technology used in digitization and formation of digital libraries. It set out some key points involved and the detailed plans required in the process, offers pieces of advice and guidance for the practicing librarians and information scientists. Digital libraries are being created today for diverse communities and in different fields e.g. education, science culture, health, governance and so on. With the availability of several free digital Library software packages at the recent time, mainly green stone digital library software is the one of them to creation and sharing of information through the digital library collections has become an attractive and feasible proposition for library and information professionals around the world.

Keywords: Digitization, Digital libraries, Librarians, Information scientists, Library software, Green stone digital library, Information professionals.

Introduction

Traditional methods of collecting, storing, processing, and accessing information have undergone a massive transformation due to the growth of virtual libraries, digital libraries, online databases, and library and information networks. Digital technology, Internet connectivity, and physical content can now be dovetailed, resulting in a digital library. Digital libraries and the digitization of print materials can preserve resources in art and culture, education, science and technology, literature and humanities, media and entertainment, and cultural heritage and history. In India, a substantial number of libraries and information centers have initiated digital library projects including databases and e-journals, or by digitizing their own archival-valuable collections. Hundreds of thousands of ancient books and manuscripts, scores of them still preserved in palm leaves, urgently need digitization to preserve the cultural heritage of India.

What are Digital Libraries?

A digital library is a collection of digital documents or objects. This definition is the dominant perception of many people of today. Nevertheless, 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 assembly of digital objects. They retain the several qualities of traditional libraries such as a defined community of users, focused collections, long-term availability, the possibility of selecting, organizing, preserving and sharing resources. The digital libraries are sometimes perceived as

institutions, though this is not as dominant as the previous definition. The following definition given by the Digital Library Federation (DLF) brings out the essence of this perception.

“Digital libraries are organization 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.”

The point in this definition is on the digital library as a dynamic, growing organism. As digital libraries evolve and become the predominant mode of access to knowledge and learning, institutionalization of digital libraries appears to be on the increase.

Requirement for Digital Libraries: The Internet and World Wide Web provide the impetus and technological environment for the development and operation of a digital library. The Internet provides the TCP/IP and or its associated protocol for accessing the information and web provide tools and technique for publishing the information over Internet. In the digital environment it is reasonable to say that a central back up or archive should be created at the national level, which will store information output of the region as well as information from outside the country. Some of the requirements for digital libraries are:

1. Audio visual: Colour T.V., V.C.R., D.V.D., Sound box, Telephone etc.
2. Computer: Server, P.C. with multimedia, U.PS. etc.
3. Network: LAN, MAN, WAN, Internet etc.
4. Printer: Laser printer, Dot matrix, Barcode printer, Digital graphic printer etc.
5. Scanner: H.P. Scan jet, flatbed, Sheet feeder, Drum scanner, Slide scanner, Microfilming scanner, Digital camera, Barcode scanner etc.

6. Storage devices: Optical storage device, CD-ROM, Jukebox etc.
7. Software: Any suitable software, which is interconnected and suitable for LAN and WAN connection. PC

Benefits of Digital Libraries

1. **No physical boundary:** The user of a digital library need not to go to the library physically, people from all over the world could gain access to the same information, as long as an Internet connection is available.
2. **Round the Clock Availability:** Digital libraries can be accessed at any time, 24 hours a day and 365 days of the year
3. **Multiple accesses:** The same resources can be used at the same time by a number of users.
4. **Structured Approach:** Digital library provides access to much richer content in a more structured manner i.e. we can easily move from the catalog to the particular book then to a particular chapter and so on.
5. **Information Retrieval:** The user is able to use any search term bellowing to the word or phrase of the entire collection. Digital library will provide very user friendly interfaces, giving click able access to its resources.
6. **Preservation and Conservation:** An exact copy of the original can be made any number of times without any degradation in quality.
7. **Space:** Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them. When the library had no space for extension digitization is the only solution.
8. **Networking:** A particular digital library can provide the link to any other resources of other digital library very easily thus a seamlessly integrated resource sharing can be achieved.
9. **Cost:** The cost of maintaining a digital library is much lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintains, rent, and additional books. Digital libraries do away with these fees.

Functional Components of Digital Library

Most digital libraries share common functional components. These include:

Selection and Acquisition: The typical processes covered in this component include the selection of documents to be added, the subscription of database and the digitization or conversion of documents to an appropriate digital form.

Organization: The key process involved in this component is the assignment of the metadata

(bibliographic information) to each document being added to the collection.

Indexing and Storage: This component carries out the indexing and storage of documents and metadata for efficient search and retrieval.

Search and Retrieval: This is the digital library interface used by the end users to browse, search, retrieve and view the contents of the digital library. It is typically presented to the users as Hyper-Text Mark-up Language (HTML) page.

These mentioned components are the important characteristic of digital library, which differ it from others collections of online information

Digitization: Witten and David (2003) defined Digitization as the process of taking traditional library materials that are in form of books and papers and converting them to the electronic form where they can be stored and manipulated by a computer.

Ding, Choo Ming (2000) has elaborated the works of Getz (1997), Line (1996) and Mckinley (1997) on the advantages of digitization. They maintained that:

1. Digitization means no new buildings are required; information sharing can be enhanced and redundancy of collections reduced.
2. Digitization leads to the development of Internet in digitalized based libraries. As Internet is now the preferred form of publication and dissemination.
3. Digital materials can be sorted, transmitted and retrieved easily and quickly.
4. 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.
5. 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.

1. To preserve the Documents: That is to allow people to read older or unique documents without damage to the originals.
2. 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 remotelocations, example, more than one person at a time.
3. 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.

Digitizing documents can take a lot of time, effort and money. Smith (2001), narrated the following reasons that should be considered before going into digitization.

Reasons to be Considered

Is it worth digitizing?

Do the documents contain the information that is valuable enough to warrant the costs of digitization? There is no point digitizing the documents that are already out of date, no matter how bulky they, but it is worthy to digitize the old, unique documents that can be easily damaged so that the people can be allowed to use them without handling the originals. These unique documents are sometimes called the heritage documents.

Who is your audience?

If there are only few users, or maybe there are large numbers of potential users, but they do not have computers to access the digital library, they can be served by sending them photocopies. It may be difficult to judge the demand for documents. It is, however; wise to get other people's opinions. Ask the potential users of the documents what they see as their priorities.

Do the documents form a collection?

It is important to verify if the documents form a collection. In fact, the documents in a digital library should have something in common like a common subject focus

How easy is it to digitize documents?

Another important factor to take into account is how easy it will be to digitize the documents. Not all the hard copy documents can be easily converted to electronic format. There is the need to check the physical characteristics of the documents to understand how easy it will be to digitize them. If you have a lot of documents that are hard to digitize, you might choose not to include them in the digital library. It is advisable to put them in the image files, rather than in the searchable text document.

According to Maxine (2000), creating a digital library collection involves the following steps: planning, implementation and promotion. These are essential if the finished product is to successfully meet the user's needs and conform with the accepted quality standards.

Planning: Planning mainly involves identifying various tasks related to creating a digital library collection, developing strategies for handling these tasks, identifying required resources and formulating a timeline for accomplishing these tasks. If there is a need to have a large digital project, you may consider conducting a feasibility study to assess the viability of the project before detailed planning. The outcome of the feasibility study could be a formal proposal for obtaining management approval or grant for the project.

- a. The first step in planning a digital library collection development project is to specify the need for creating the digital library collection, its purpose and target user community. You should indicate if management, the users or others have expressed this need and defined what this need is. The purpose could be improving preservation of some rare or delicate materials, improving access to and the visibility of certain material or facilitating re-use of documents. It is important to identify the target user community for a digital library collection and their profile.
- b. There is the need to define the source material that constitutes the digital library collections and the key attributes of this source material. Examples of source material include project reports, staff publications, working papers, theses, dissertation, audio and video lectures, songs and musical scores etc. There is also the need to specify what portion of the material is to be digitized and if all the material or only a sub-set will be covered in the digital collection. Remember to assess copyright restrictions.
- c. Define the key features of the digital library collection you plan to build. Identify the nature of the collection e.g. static or dynamic. Indicate the type of usages you would allow the users to adhere to and the kind of service delivery they should expect from you e.g. CDROM or on-line or both. Define metadata, search and retrieval requirements.
- d. The important task in creating a digital library collection is the conversion of the source materials available in hardcopy into a digital format. There should be a clear cut statement about the related requirements and their processes, namely:
 1. How to convert the source material into required digital format.
 2. What are the digitization requirements?
 3. The workflow involved in digitizing the source material.
- e. Identify the resources and money required for creating and maintaining digital collections. There is a need to identify:
 1. What type of information technology (IT) infrastructure is required for establishing and maintaining the digital collections?
 2. What are the personnel requirements and
 3. What are the financial requirements involve for setting up and maintaining the collection.
- f. Finally, there is the need to define how the project is going to be implemented and what the major milestones and time requirements are?

Implementation: Planning is followed by implementation. That is getting down to the actual steps required to set up the collection. This means that there must be a need to obtain the management approval for

the plan and the required resources before proceeding with the implementation.

There is a need to identify and designate a project manager to lead the implementation of the digital project. For large digital library projects, it is essential to have a full time project manager for the project period. The Implementation of a digital library project involves the following activities.

1. Establish the project team
2. Set up the Information Technology (IT) infrastructure
3. Procure and install digital library software
4. Finalize policies and specifications
5. Complete arrangement of workflow for digitization
6. Set up the digital library collection site in case of Internet distribution
7. Obtain copyright permissions and
8. Release the digital library collection for use.

Promotion and Provision of Services: The digital library collection created should be visible, and it should provide an easy access for users. One-way of achieving this is to include links to the collection site in the appropriate pages of the library website and other related on-line services in the organization.

In addition to, or in the absence of remote on-line access to the digital collection, there is the need to explore other modes of providing access to the digital collection. These may include:

1. Setting up local public access computers on the library Local Area Network.
2. Provision of e-mail based services and
3. CD-ROM based distribution of the collection.

Technology Infrastructure and Personnel: Several resources are required for the creation of digital library collections, their maintenance and provision of services. The two major resources needed are technology infrastructure and personnel.

Infrastructure: Access to a digital library collection can be provided on-line or off-line. The On-line access today typically means that the client uses a web browser on a desktop computer or laptop and access the collection by connecting to the digital library website over the Internet. The On-line access requires a connection to the Internet or to an internal network (Intranet). In Off-line access, the digital library is not accessible over a network. One way of providing an Off-line access to a digital library collection is to receive and respond to the user queries over e-mail. Another way is to distribute the digital library collection on a CD-ROM.

A digital library project would typically require the following equipment: Server computer, Desktop computers, Digitization equipment, Network connectivity and other equipment. Another aspect is the software to be used in digital library. The Digital library software works with the web server in providing various digital library functionalities including creation, organization, maintenance, indexing, search and

retrieval. In choosing the software, some features should be taken into consideration. These include: Support for different document types, Support for customized metadata, Collection administration, Support for standards like Dublin core metadata standard, Search and retrieval and Multi-lingual support.

Several free digital library software packages are now available which could facilitate the easy creation and sharing of information through digital library collections. Examples of open source free digital library software include: Greenstone Digital Library software by New Zealand Digital Library; Academic Research in the Netherlands On-line (ARND); Tilburg University, The Netherlands; CDS ware; CERN Document server software, Geneva, Switzerland; D-space; MIT Libraries, Cambridge, MA USA. Etc.

Personnel: Personnel are most important digital library's resource, not only during its initial creation and set up, but also for its operation, maintenance and provision of services. Since the access to the digital library is easy, compared to a physical library, more users are likely to access it. If the digital library does not meet the expectations of the users in terms of currency and quality of content, they will lose confidence, and it is likely for them not to visit the digital library again. It is therefore important to assign the personnel with the right skills and attitude to handle the various tasks associated with the digital library project. Broadly speaking, the personnel will be required for the following tasks:

- a. Project management
 1. Selection and preparation of source material
 2. Digitization and conversion
 3. Cataloguing and metadata assignment
 4. Quality assessment
 5. System administration and maintenance of digital library server and website.
 6. System analysis/programming for digital library application/interface development
 7. Promotion and provisions of services.

Moreover, the rapid changes in the digital library technologies require constant re-training and re-positioning of staff for an effective practice in technological application.

Greenstone Digital Library Software: Greenstone is a freely available suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet or on the CD-ROM. The Greenstone is open source software, issued under the terms of the GNU General Public License. The aim of the software is to empower the users, particularly in the Universities, Libraries and other public service institutions, to build digital libraries. The software has the following features such as multi-platform availability for windows, Linux, access and distributed through the Internet, Intranet and CD-ROM, powerful indexing from full-text and

creation of indexes for various metadata, powerful search and browse, support different file formats (html, pdf, doc rtf, ppt etc), extensibility by allowing customization and configuration. Greenstone also allows the building of non-textual multimedia such as audio, video and pictures accompanied by textual description to allow for searching and browsing.

Installation Greenstone Digital Library Software Download Sites

The GSDL 2.72 software can be downloaded from

<http://www.greenstone.org/or>

http://greenstonesupport.iimk.ac.in/our_mission.htm.

Java2 Runtime Environment can be downloaded from <http://java.sun.com/>

Image Magick can be had from

<http://www.imagemagick.org/>

Ghostscript can be obtained from

<http://www.cs.wisc.edu/~ghost/>

Installation Procedure of Software: Install Java2 Runtime Environment in your computer as a pre-requisite for installing GSDL software. Then install the GSDL 2.72 (windows version) software in your computer. Choose the Local Library mode for installation. You may install ImageMagick and Ghostscript software's, required to build image collections and to do advanced conversion of PDF and Postscript documents, respectively. While installing the software's, just choose the default options in the installation wizard.

Building up a Collection with GLI: The simplest way to build a new digital library collection is to use *Greenstone's Librarian Interface1 (GLI)*, a component of Greenstone Digital Library software. GLI allows one to collect the documents, import or assign metadata, and build the documents into a digital library, and convert it into a CD-ROM library.

GLI can be used to perform the following basic activities, while building up a collection: -

1. **Gather** documents for building up the collection
2. **Enrich** the documents by adding metadata
3. **Design** the collection, its appearance and the access facilities
4. **Format** the appearance of the digital library
5. **Create** the collection
6. **Convert** the digital library into a CD-ROM library

Conclusion

Digitization has opened up new audiences and services for libraries, and it needs to be integrated into the plans and policies of any institution to maximize its effectiveness. Digitization is a complex process with many crucial dependencies between different stages over time. Utilizing a holistic life-cycle approach for digitization initiatives will help develop sustainable and successful project.

A general-purpose digital library like greenstone is a useful tool to provide information services in our libraries. Absence of knowledge on how to use it should

not come in the way of exploiting the advantages it offers. This documentation may be used as a tool to bring in more people to the growing constituency of greenstone users. We, the librarians, can improve our capabilities, as knowledge managers, if we are particular in learning the information technology tools like greenstone and use them for managing knowledge resources. We should learn, utilize, promote and propagate greenstone to make our libraries better.

References

1. Alhaji, I. U. (n.d.). Digitization of library resources and the formation of digital libraries: A practical approach. 1-17. Retrieved February 2, 2015, from <http://netlibrarynigeria.net/downloads/Ibrahim's%20paper.pdf>
2. Chowdhury, G. (2010). From digital libraries to digital preservation research: the importance of users and context. *Journal of Documentation*, 66 (2), 207-223. Retrieved February 7, 2015, from <https://comminfo.rutgers.edu/~tefko/Courses/Zadar/Readings/Chowdhury%20dig%20pres%20J%20Doc%202009.pdf>
3. Ian H. Witten, D. B. (2006). Digital libraries in education. MOSCOW: UNESCO Institute for Information Technologies in Education (IITE). Retrieved February 3, 2015, from <http://iite.unesco.org/pics/publications/en/files/3214563.pdf>
4. N. Varatharajan, M. C. (2007). Digital Library Initiatives at Higher Education and Research. *Library Philosophy and Practice*, 1-7. Retrieved February 5, 2015, from <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1158&context=libphilprac>
5. Pavani, A. M. (2007). The Role of Digital Libraries in Higher Education. *International Conference on Engineering Education*, 1-7. Retrieved February 9, 2015, from <http://www.ineer.org/events/icee2007/papers/637.pdf>
6. RAJASEKHARAN, K. M. (2007). Building up a Digital Library with Greenstone A Self-Instructional Guide for Beginner's. Mulagunnathu Kavu, Thrissur: Kerala Institute of Local Administration. Retrieved February 10, 2015, from <http://eprints.rclis.org/9139/>
7. Tutorial exercises on Greenstone Retrieved February 7, 2015, from http://greenstone.sourceforge.net/wiki/index.php/Tutorial_exercises
8. Vrana, R. (n.d.). The Importance of Usability in Development of Digital Libraries. *Stručni članak*, 379-387. Retrieved February 7, 2015, from <http://infoz.ffzg.hr/INFuture/2007/pdf/5-01%20Vrana,%20Radovan,%20The%20Importance%20of%20Usability%20in%20Development%20of%20DL.pdf>