# DIGITIZATION OF LIBRARY RESOURCES AND THE FORMATION OF DIGITAL LIBRARIES: AHMADU BELLO UNIVERSITY'S EXPERIENCES

By

Ibrahim Usman Alhaji
E-mail: uscoma@yahoo.co.uk
Mobile Phone: +234-803-592-0491

**Abstract:**
*Digital Libraries are being created today Nigeria for diverse communities and in different fields e.g. education, science, culture, development, health, governance and so on. With the availability of several free digital Library software packages at recent time, the 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. This paper examines the activities, methods and technology used in digitization and formation of digital libraries at Kashim Ibrahim Library, Ahmadu Bello University, Zaria. It set out some key points involved and the detailed plans required in the processes, offers pieces of advice and guidance for the practicing Librarians and Information scientists. The paper ends with a call to integrate digitization into the plans and policies of any institution to maximize its effectiveness.*

**Keyword: Digitization, Digital Library, Greenstone software,**

## Introduction

The availability of several free digital library software packages (Greenstone Digital Library Software) at recent time, the formation of digital library and sharing of information has become an attractive and feasible proposition for libraries and other institutions around the world. Although Library automation has helped to provide easy access to bibliographic data through the use of computerized library catalogue such as On-line Public Access Catalog (OPAC). Digital libraries differ significantly from the traditional libraries operation because they allow users to gain an on-line access to and work with the electronic versions of full text documents and their associated images. Many digital libraries also provide an access to other multi-media content like audio and video.

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 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 and possibility of selecting, organizing, preserving and sharing of 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." (DLF, 2001)

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 is therefore necessary.

On the other hand digitization as defined by Witten and David, (2003) is 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:

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 mentioned, 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. Some of the advantages of digitization are to preserve materials in order word to allow people to read, listing or watch older or unique resource without damage to the originals. Also to make the resource more accessible by serving the existing users better and remotely

Digital Libraries are being created today Nigeria for diverse communities and in different fields e.g. education, science, culture, development, health, governance and so on. For an institution to form or create a successful digital library there rules and guideline that need to be follow. Maxine, (2000) affirmed that creating a digital library collection involves some steps such as planning, implementation and promotion. These are essential if the finished product is to successfully meet the user's needs and conform to the accepted quality standards. The National Center at Kashim Ibrahim Library Ahmadu Bello University, Zaria, Nigeria adopted the processes as outlined by Maxine, (2000).

**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. 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:

i. How to convert the source material into required digital format.
ii. What are the digitization requirements?
iii. The workflow involved in digitizing the source material
iv. Identify the resources and money required for creating and maintaining digital collections.
v. What type of information technology (IT) infrastructure is required for establishing and maintaining the digital collections?
vi. What are the personnel requirements and
vii. What are the financial requirements involve for setting up and maintaining the collection.

**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:

i. Establish the project team
ii. Set up the Information Technology (IT) infrastructure
iii. Procure and install digital library software
iv. Finalize policies and specifications
v. Complete arrangement of workflow for digitization
vi. Set up the digital library collection site in case of Internet distribution
vii. Obtain copyright permissions and
viii. Release the digital library collection for use.

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.

The center however, has structured its digitization stages from registering, scanning, proofreading, formatting and conversion to portable document format (PDF), building collection, application of metadata and final uploading to the library website.

i. **Registering**

Scanning large number of documents is confessing and difficult. 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 sub-directory. Moreover, a good filing system is vital, so everyone in the digitizing team knows what he supposes to do, and he can fill in for another person in case of absence. The center takes this process as an important and first step in digitizing its resources.

ii. **Scanning**

The center have acquired Xerox Doc-mate 750 an automatic document sheet feeder machine that scan 100 pages of thesis and dissertation in few minutes. The process is simple and easy when the machine software is set to a define resolution and colour. After cleaning and dusting off the documents to be scanned. The Center make sure that all the pages are present and in the right order. If the document is in poor condition, the center tries to find a fresh copy. The center cut the document open to get individual sheets to feed through the scanner and send document for rebinding later. However, documents that the center does not want to damage, it photocopy each page and feed in the photocopies through the scanner. This process also applies to Newspapers and other resources slatted for digitization.

However, to scan a document on a flatbed scanner, the document can be placed face down on the scanning machine and instruct scanning process from the

machine software. Moreover, the center does not engaged much in Optical Character Recognition (OCR) scanning as much of the resources are direct image scanning.

### iii. Proofreading, formatting and conversion to portable document format (PDF)

This is the act of making corrections to the document text and layout, however, since the center does not engaged much in Optical Character Recognition (OCR) scanning as much of the resources are direct image scanning what takes much time  of the staff is the conversion to portable document format (PDF). This means converting the image file to portable document format (PDF) through the software called Adobe converted.

Nevertheless, for resources that are born digital proofreading and formatting has become necessary. The center makes corrections to the document text and layout before converting.

### iv. Building collection

**Starting a New Collection**
- Start a new collection within the Librarian Interface:
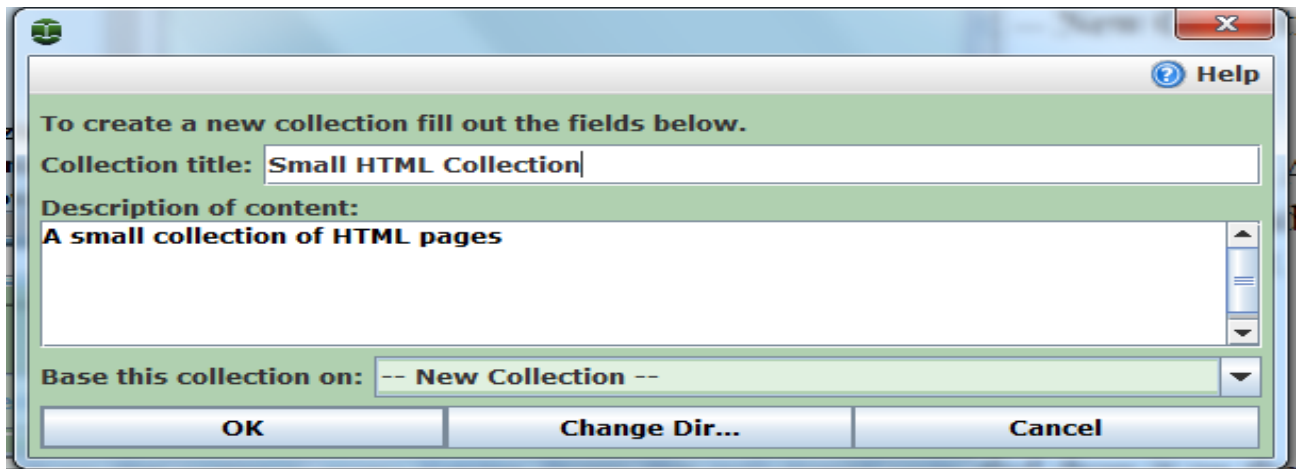  **File → New...**
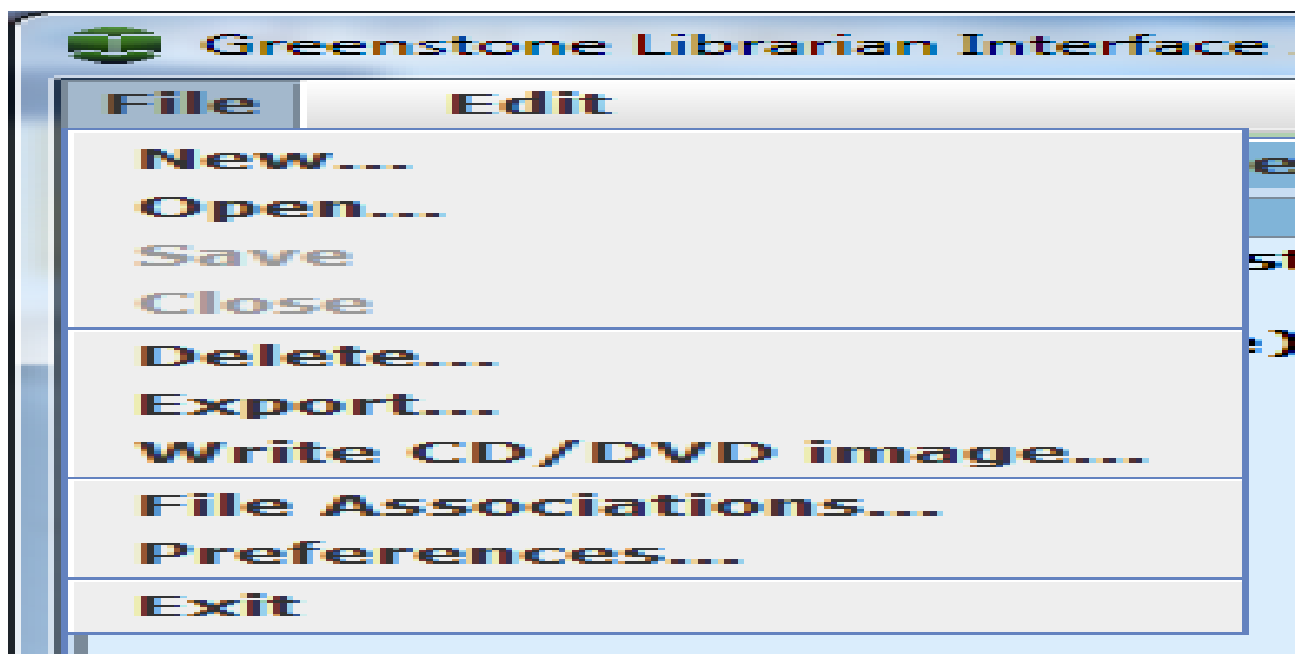
- You will create a collection based on scanned pages.

A window pops up. Fill it out with appropriate values—for example,
**Collection title:**    Small HTML Collection
**Description of content:** A collection of Thesis. Leave the setting for **Base this collection on:** at its default: **-- New Collection --**, and click **<OK>**.



- Next, you will need to gather together the files that will constitute the collection. A suitable set has been prepared ahead of time that is scanned thesis. Using the left-hand side of the Librarian Interface's **Gather** panel, interactively navigate to the folder.

## Adding documents to the Collection

- Now drag the *scanned convertible document's* folder from the left-hand side and drop it on the right. The progress bar at the bottom shows some activity. Gradually, duplicates of all the files will appear in the collection panel.

  *You can inspect the files that have been copied by double-clicking on the folder in the right-hand side.*
- Since this is our first collection, we won't complicate matters by manually assigning metadata or altering the collection's design. Instead we rely on default behaviour. So pass directly to the **Create** panel by clicking its tab.

## Building the Collection

- To start building the collection, click on **Create** then click the **<Build Collection>** button.

- Once the collection has been successfully built, a window pops up **"Collection Creation Result"** windows pop-up will appear. Click **<OK>** to confirm the creation.
- Click the **<Preview Collection>** button to look at the end result. This loads the relevant page into your web browser (starting it up if necessary).

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.

v. **Application of metadata**

**Adding Title and Description Metadata**
- We work with just the first three files to get a flavour of what is possible. First, set each file's **dc.Title** field to be the same as its filename but without the filename extension:

  Click on *Thesis title* so its metadata fields are available, then click on its **dc.Title** field on the right hand side. Type in **the title of the thesis**

  Repeat the process for other titles.

- Add a description for each image as **dc. Description** metadata.
  What description should you enter? To remind yourself of a file's content, the Librarian Interface lets you open files by double-clicking them. It launches the appropriate application based on the filename extension, Word for .doc files, Acrobat for .pdf files and so on.

  Double-click *the thesis* on Windows, the image will normally be displayed by Microsoft's Photo Editor (although this depends on how your computer has been set up).

  Back in the **Enrich** panel; make sure that *thesis title* is selected in the collection tree on the left hand side. Enter the text I e the thesis title as the value for the **dc. Description** field.

  Repeat this process for the others, adding a suitable description for each.

- Go to the **Create** panel and click **<Build Collection>**. Once it has finished building, **preview** the collection. You will not notice anything new that is because we haven't changed the design of the collection to take advantage of the new metadata.

**Change Format Features to display new metadata**

- Now we customize the collection's appearance. Go to the **Format** panel and select **Format Features** from the left-hand list. Leave the feature selection controls at their default values, so that **All Features** is selected for **Choose Feature**, and **VList** is selected as the **Affected Component**. In the **HTML Format String**, edit the text as follows:
  Change _Image Name_: to Title:
  Change [Image] to [dc. Title]
  After [dc.Title]<br> add Description: [dc.Description]<br>

*Metadata names are case-sensitive in Greenstone: it is important that you capitalize "Title" and "Description" (and don't capitalize "dc").*

- The new format statement is displayed in the list of assigned format statements. The first substitution alters the fragment of text that appears to the right of the thumbnail image; the second alters the item of metadata that follows it. The addition displays the description after the Title.

- Preview the collection by clicking the **<Preview Collection>** button. When you click on **Browse** in the navigation bar the presentation has changed to "Title: Bear" and so on. Each image's description should appear beside the thumbnail, following the title.

vi. **Final uploading to the library website**

To publish a collection on Website, CD-ROM or DVD, Greenstone's Export to CD-ROM export module must be installed. This is included with CD-ROM distributions and all distributions 2.70w and later. It must be installed separately for non-CD-ROM versions of Greenstone, version 2.70 and earlier (see **Installing Greenstone**).
1. Launch the Greenstone Librarian Interface if it is not already running.
2. Choose **File → Write CD/DVD or website image...**. In the resulting popup window, select the collection or collections that you wish to export by ticking

their check boxes. You can optionally enter a name for the CD-ROM: this is the name that will appear in the menu when the CD-ROM is run. If a name is not entered, the default **Greenstone Collections** will be used. You can also specify whether the resulting CD-ROM will install files onto the host machine when used or not. Click **<Write CD/DVD image>** to start the export process. The necessary files for export are written to: Greenstone → tmp → exported_xxx where xxx will be similar to the name you have entered. If you didn't specify a name for the CDROM, then the folder name will be exported collections.

## Some good examples of collection in CD
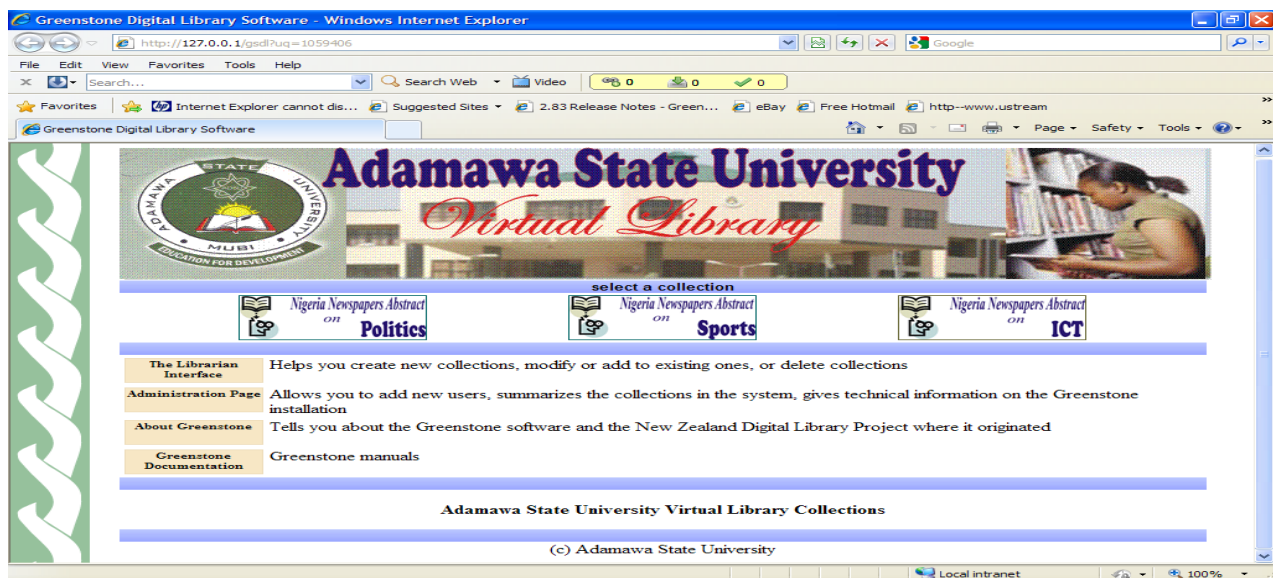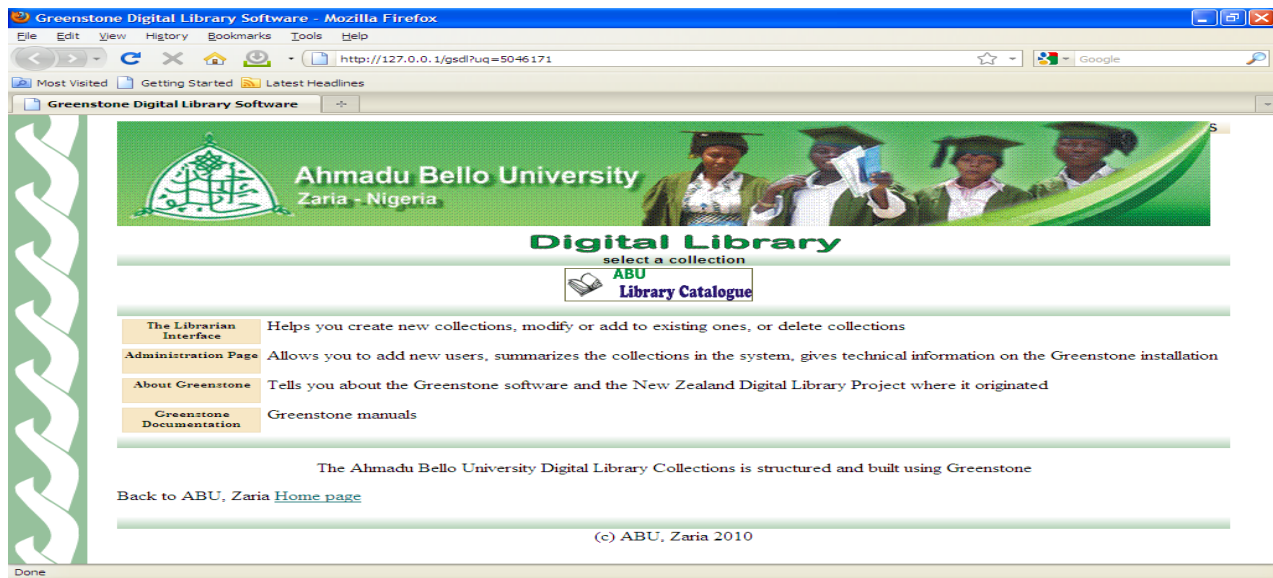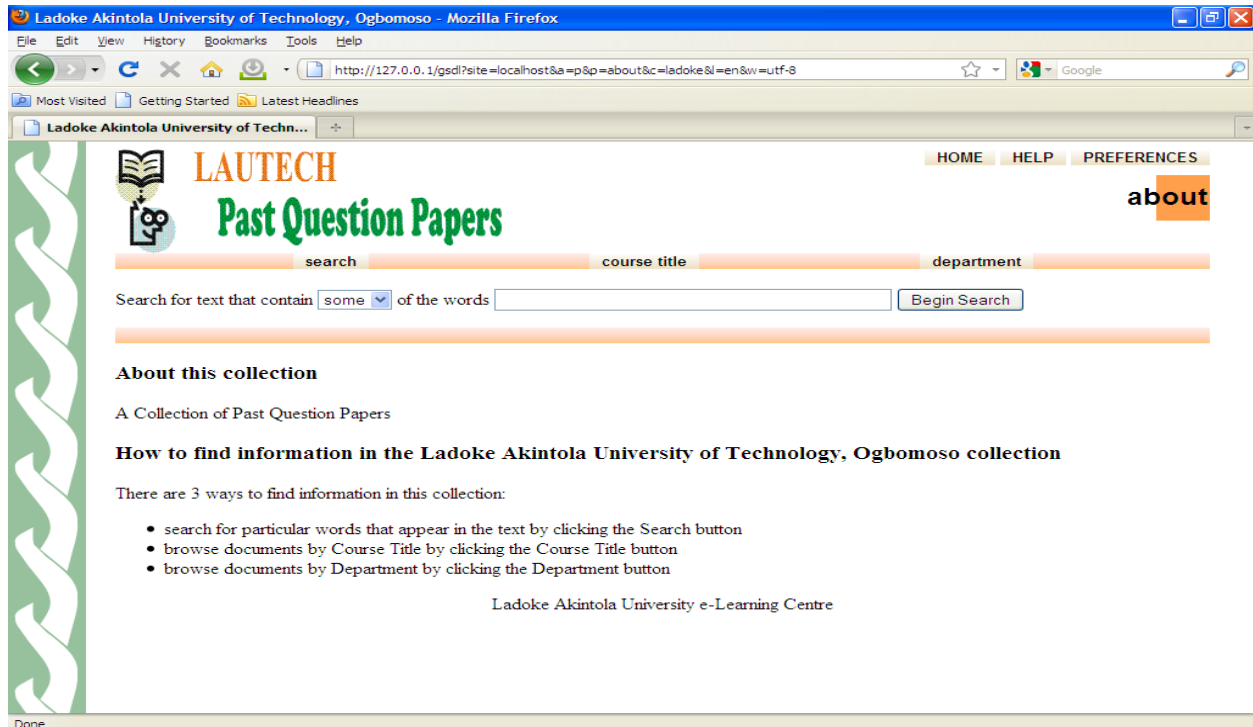
## Infrastructure

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.

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.

In its efforts to have a clean digitized resources the center have acquired Xerox Doc-mate 750 an automatic document sheet feeder machine that scan 100 pages of theses and dissertation in few minutes

**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 organization website and other related on-line services in the institution.

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:

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

**Conclusion**

Digitization and digital library 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.

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 and building a reboots digital library that will stand the text of time.

## Reference

1. Cornell University Library/Research Departments (2000), *Moving theory into Practice: digital Image for libraries and archives.* Research Libraries Group. Available at http://www.library.cornell.edu/preservation/tutorial

2. Digital Library Federation. (2001), Registry of Digitized Books and Serial Publication, Available at http://www.digilib.org/collections

3. Ding, Choo Ming. (2000), Access to Digital Information: Some Breakthrough and Obstacles, Journal of Librarianship and Information Science, Vol.32 No.1

4. Greenstone Training Workshop Material.(2002), Available at http://www.greenstone.org

5. Ian, H. Witten & David, Bainbridge. (2003), *How to Build a Digital Library,* London: Morgan Kaufman Publishers

6. Sitts, Maxine K. (2000), *Handbook for Digital Projects: A Management Tool for Preservation and Access.* Northeast Document Conservation Center, Andover, Massachusetts. USA. http://www.nedcc.org/digital/dman.pdf

7. Smith, Abbey (2001), *Strategies for Building Digitized Collection.* Washington, D.C. Digital Library Federation, Council on Library