Abstract

Digital library is nothing but the digital version of a library. In other words, it is a place where all digital documents are stored or collected. In other words, it is an organised as well as focused collection of books, journals, periodicals, audio visual materials, which can be accessed and retrieved. This paper introduces digital library and Metadata Service Providers. It also throws light on metadata harvesting using D-space taking Dyuthi repository as an example. The terms like network of technology, information, documents, people and practices get interconnected together and give rise to a new word known as Digital library. Today digital libraries are increasing at a higher, with lots of choices and option and more even day to day users are increasing which over all show the total usage of people for digital libraries.

1. Introduction

Libraries are obviously a place where we visit, in all phases of our lives. Right from a child, to a senior citizen, there is no criteria for using the library. Throughout the years, many things have changed, both inside as well as outside libraries. The catalogue cards, the membership cards, the books, the journals, the periodicals. What more to say, the whole thing of conventional libraries hardly exist now. And voila, here we libraries right in front of our eyes, on a screen, in your homes. And that is where technology has taken us. We sit in the rooms and access records published in
nooks and corners of the world. Let us take a tour and observe those little magic happening around us. This paper introduces the virtual library, digital library and electronic library. It also throws light on metadata harvesting using D-space taking Dyuthi repository as an example.

2. Functions Of A Digital Library In Digital Environment

Like any other library, digital library has its own functions. There are much similar to that of the normal libraries.

I. Selection and acquisition
This generally includes the selection and procurement of documents. It also includes the conversion of the available documents to a digitised form.

II. Organisation
This includes assigning the metadata to the existing documents, so as to make retrieval easy.

III. Indexing and storage
This helps in organising the data in a systematic manner so as for easy retrieval and effective use of data.

IV. Search and retrieval
This is made happen by creating a user interface using Hyper Text Markup Language. It helps in searching and retrieving the documents needed by the users.
- It enables discovery of new application for old knowledge.
- It brings prestige to the person and the institution.
- It promotes progress of the society.

3. Bridging Digital Divide

ICT can be effectively used for bridging the digital divide. The following areas are very much important to have bridging the digital divide:

I. Physical connectivity – by providing road
II. Electronic connectivity – by providing reliable communication network.
III. Knowledge connectivity – by establishing professional institutions and vocational training
IV. Economic connectivity – to enhance the prosperity of clusters of libraries in rural areas, which is to be brought about in a most creative and cost-effective manner. (Sinha, 2006)

3.1: Significance of digital libraries

- They help preserve the original documents
- Helps the documents to be reusable
- Helps in making it accessible to a larger audience

3.2: Component of Digital Library

3.2.1: Collection Infrastructure
i. Acquiring of collection available in Digital formats.
ii. Buying access to external Digital Collections i.e. Online Databases (EBSCO, Emerald and Sage etc).
iii. Conversion of Existing Print Media into Digital Format means Digitization of available important resources.
iv. Print to Digital: Options for conversion: It involves steps as Scanned as image, OCR and retaining layout using software and re-keying the data.
v. Steps in the process of Digitization i.e. pre, actual and post procedure.
vi. Formats and Encoding used for text as ASCII, unstructured text and structured text(SGML, HTML or XML)
vii. Converting PostScript files to PDF files i.e. Format TIFF (Tagged Image file Format), BMP (Bit Map Page) and PNG (Portable Network Graphic) etc.

viii. Organizing Digital Projects: It covers indexing of objects which could digitize.


tax. Providing integrate access interface of e-resources to user.

3.2.2: Access Infrastructure

i. Search and browsing interfaces creates with simple graphics which will understand by user.

ii. Including Information Retrieval with basic filter for better search in Digital Library.

iii. Provide links of Meta Resources, Portals and Knowledge Gateways.

3.2.3: Computer and Networking Infrastructure

- Networking: digital Library can build on LAN, MAN and WAN network but set up in user convenient approach access. It gives different environment for different networks. Mostly LAN is used for digital library. It is used on intranet

- Set main or additional Server which help in store digital objects which could digitize and subscribed by library.

- Input devices requires like scanners, digital cameras, video comers with high resolution pixel.

- Storage devices: external hard-disk, portable devices and server for back-up.

- Communication devices: It requires digital library software like Greenstone, E-prints, Fedora and Ganesha.

- Image capturing or scanning software: OminiDoc ver.1(Newgen Software) and Data Scan (Stacks Software Pvt. Ltd).


- Web service: setting –up in different environment like Unix, Windows and Linux.

- Optical Character Reorganization (OCR) use for capturing character appears in image. Text Bridge (Scansoft) and OminiPage (caere).

- Database Management Software (DBMS): SQL (Structured Query language) are used.

- Client-side-Hardware and Software Components: Need Internet Browser, Acrobat software package and any PDF viewer,

- Digital resource organization

- Object naming and addressing: Location of objects can determine based on URL, PURLs (Persistent Uniform Resource Locators), URN (Uniform Resource Name) and DOI (Digital Object Identifier).

- Online Database connectivity (ODBC) driver requires installing in computer.

- Uniform Resource Characteristics.

3.2.4: Advantages

Digital libraries as like any other technology, has a lot of benefits. A few of them are listed below:

- Improved Access: since they are accessed through CD’s or internet, it can be accessed from anywhere at any time. It is very portable and compact.

- Better coverage: since they are soft copies, multiple copies can be given making many users to access it at the same time.

- Preservation: since the original documents are not directly accessed, the original are always safe and undamaged.

- Resource sharing: it can easily make many users use easily. (Aljahi)
3.2.5: Promote use of Digital Library

- **Web-Service**: library web-site is mirror of library. It gives new service like RSS feed for automatically upload latest news. As well as portal links which can easily visible to user. Library can design web-site in approach of user friendly.

- **Band-width Speed**: Organization/ government gives high-speed band-width which is helpful to user in accessing online resources like 4G and 5G speed. It provides faster access to resources.

- **Advanced Search Facility**: In Digital Library no need of cataloguing and classified of resources. Just library must to add advanced search tool with help of Boolean operators, year-wise search, full text and many filters. If search facility is good then user can reach to his/her resources.

- **Full-text Content**: User can watch full-text content and download the content. Library staff is required to set permission for full text in Digital Library Software. It is helpful to user to access whole information at remote area.

- **User-friendly Visualization**: Site visualization and customization facility can create which attracts to user / user-friendly. It helps in increasing use of digital library world wide web.

- **Metadata Standard**: Initial Stage organization/ institute / government firstly decide staff understandable format for harvesting of metadata. Many library referred Dublin Core Metadata Standards. It help n describing objects on web or data which describing and discover of content / resources.

- **Easy Mechanism**: Library provides an easy mechanism for resource sharing with other libraries and user. As well as library can make links on share on Social Media. **Decisions about discoverability services are practical only to the extent that library resources can sustain these efforts.**

3.2.6 Metadata

- Descriptive metadata should be delivered with all necessary contextual information,

- Access points should include a uniform data string for retrieval, such as a resolvable URI or a text string

- Both the content and format of metadata should be kept up to date in the home repository, and updated metadata should be made available or pushed to external sites.

- Metadata should adhere to shared standards, in order to facilitate and simplify the data transformations necessary for successful metadata exchange and interoperability.

4. OAI/PMH (Open Archive Initiative/ Protocol Metadata Harvesting)

Before moving on to OAI/PMH let us see what a metadata is or what is meant by metadata. Metadata in simple word is nothing but data about data. Now being more specific, metadata is anything that describes an information resource in a well-structured manner. Whether it be any data used to access, locate, or retrieve information, it all comes under metadata.

There are three features of metadata:

- **Context**: it is basically on how, when, where and more. It gives to the context of the information.

- **Content**: is generally about what is inside.

- **Structure**: is based on the associations of the information sets.

The above are the basic things about metadata. Now taking about OAI/PMH, its nothing but a harvesting tool. It is used to harvest metadata as well as full text data from different repositories. It uses Hyper Text Transport Protocol and extensible markup language for harvesting techniques.
4.1 Why OAI/PMH?
With this software, we are able to build up a connection between different academic communities throughout the world. It helps in resource sharing and and makes ir accessible and reachable for a larger audience.
Building open one’s own repository encourages the growth of research and development and also leaves a lot of reference articles for the upcoming generations to study on.

4.2 Functions of verbs in OAI/PMH
Let us know more about OAI/PMH through by taking D-space as an example. D-space is an open source repository software used to create open access repositories where scholarly articles and journals are published. It was released on Nov, 2002 by a group effort by MIT and HP Labs.
There are generally six verbs, each with different functions.
- GetRecord
- ListSets
- ListRecords
- ListIdentifiers
- ListMetadataFormats
- Identify

Where Record is metadata in a designed format item is any record, collection or component from the repository, identifier is a unique key in the repository, set is something that groups the items in the repository.
Identify
This generally tells about the capabilities of the repository. It basically describes the repository. It’s much like how we introduce ourselves. It is the identity of the particular repository.

Figure 1: Dyuthi - the repository of Cochin University of Science and Technology
How the request is formed:
rooturl/oai/request?verb=Identify

**Repository (Example Figure Shown Above)**
Eg: take the repository of Cochin University of Science and Technology (CUSAT). Dyuthi is the name of the repository.
dyuthi.cusat.ac.in/oai/request?verb=Identify

**ListMetadataFormats**
This lists out the metadata formats supported by the repository.
How the request is formed:
rooturl/oai/request?verb=ListMetadataFormats

---

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
<OAIPHMetadata xmlns:oaipath="http://www.openarchives.org/OAI/2.0/" xmlns:oai="http://www.openarchives.org/OAI/2.0/"
    oai_dc rdf:about="http://www.openarchives.org/OAI/2.0/oai_dc">
    <responseDate>2014-09-17T04:44:01Z</responseDate>
    <metadataCollection>http://dyuthi.cusat.ac.in/oai</metadataCollection>
    <metadataFormat>http://dyuthi.cusat.ac.in/oai/request?verb=Identify</metadataFormat>
    <metadataFormat>http://dyuthi.cusat.ac.in/oai/request?verb=ListMetadataFormats</metadataFormat>
    <metadataFormat>http://dyuthi.cusat.ac.in/oai/request?verb=ListSets</metadataFormat>
</OAIPHMetadata>
```

---

**ListSets**
This lists all the collections presented in the repository. Every collection has a unique number called the handle number through which they are harvested.
How the request is formed:
rooturl/oai/request?verb=ListMetadataFormats
eg: dyuthi.cusat.ac.in/oai/request?verb=ListSets
Figure 3: Example- dyuthi.cusat.ac.in/oai/request?verb=ListSets

Here as we can see, hdl_purl_942 harvest all the files in that collection named APT 2008.

**ListIdentifiers**

This gives the OAI unique id that is contained in the repository.

How the request is formed:

rooturl/oai/request?verb=ListIdentifiers

eg: dyuthi.cusat.ac.in/oai/request?verb=ListIdentifiers

Figure 4: Example- dyuthi.cusat.ac.in/oai/request?verb=ListIdentifiers
5. How to harvest metadata?

The metadata is harvested in different software’s by using different methods. In D-space, it is done by following the below mentioned processes.

- We go to edit collection
- Then we go to Content Source
- We check the column that says This collects harvests its contents from an external source.
- Fill the base Universal Resource Locator (URL) and add the request eg:dyuthi.cusat.ac.in/oai/request
- Say which set we want to harvest
- Fill in the handle number
- Check simple Dublin Core
- Say what u want to harvest
- Test settings
- Once notification comes, harvest the metadata.

Every metadata record get harvested through OAI-PMH, it consist of identifier which tells from which the metadata was derived. OAI-PMH is not a search protocol, but performs search based services. Dublin core is used as for the interoperability, OAI-PMH is easy to deploy with many toolkits that can hide the protocol from developers”.

![Figure 4: Digital Resources (Metadata and full text ) Discovery Process](image-url)
6. Advantages of OAI/PMH

OAI/PMH has a lot of advantages to be listed out.

- It helps in the creation of repositories.
- Helps in forming a network between different academic institutions, so that we can get the best out of everything.
- Helps get the information from anywhere around the world very fast and accessible and reachable to a greater audience with in stipulated time.
- Helps it to be a knowledge resource.
- Harvesting metadata has become very easier.
- Good quality of articles.

7. Limitations

Like every coin that has its two sides, this also has some limitations which are negligible compared to the advantages and the hardships are worth the benefits.

- We can’t expect it to have all documents in the world.
- Forming network of all the repositories is a tough job.
- Technical issues and risks.

But at the end of the day the work done on forming a networks and the chaos related in doing so are worth taking for as it encourages research and development which is of peak importance. Some effort from few can help save the effort of a million. A little time of yours can save that of many.

8. Conclusion

In the age of digital era, so many things come around. Present day user want information on desk and furthermore want Internet connection with high speed. The coming of digital libraries have helped a lot in making documents accessible very easily in limited period of time. The involvement of technology in the field of education and information has taken it to better heights and standards. Like the cherry on the icing, metadata harvesting and harvesting techniques have given a huge input in research and development making materials available to all, without much trouble and effort.

Present situation digital library can’t replace by physical existent of resources. Therefore library becomes hybrid nature. Day by day electronic publication increased tremendously. In future many libraries will shift towards a digital environment. Large scale digitization projects are going on like Google Books, The million Book Projects and Gutenberg project. Above all this, it saves a lot of time and energy of the users, making it available to them at the touch of their finger tips.

References