CHAPTER VI

DATA ANALYSIS AND DIGITAL LIBRARY MODELS
6.1 Introduction:
The present chapter is divided into two parts. The first part deals with the information obtained from Indian University libraries through questionnaire and DLI-2 projects by visiting the websites. The second part is devoted to discuss the digital library models. The present survey was undertaken to know the extent of digitization activities carried out in India and abroad. It is also necessary to find out the priorities, policies, best practices, funding agencies and staff requirement in the field of digitization and digital libraries such type of surveys had been carried out in the past in developed countries like USA, UK, Australia, Canada etc. Based on their experiences the researcher designed a structured questionnaire containing 31 questions. The main goal of this questionnaire was to find out what is the current situation of digitization in India. At the same time it was decided to study twenty-eight digital library projects to compare and contrast the situation between India and developed country like USA.

Hence it was decided to select all the Indian Universities and digital library projects, which were, including in the DLI-2 phase. Though there were more 235 universities, listed by AIU in 1999 handbook, it was decided to include those universities, which had e-mail facility. This selection criteria is laid on the basis that communication technology is one of important ingredients of establishing a digital library. These e-mail addresses where obtained from INFLIBNET Website. The DLI-2 projects have their own website providing comprehensive information on its history, objectives, achievement and research work etc. Which is also considered for analysis purpose.

After selecting the appropriate samples (institution), the researcher sent e-mail to 200 Indian universities and 28 DLI-projects. It was found that out of 200 e-mails 90 mails returned back, the reasons may be different. With
respect to DLI-2, five project Directors replied immediately, but they informed refer their web-sites for getting answers to these questions.

Since reply to e-mail questionnaire was very poor it was decided to send printed questionnaire to all 200 Indian University libraries and hunt information from 28 DLI-2 project's website. Response to this was also very low. Multifold efforts were made to attain maximum response from these organizations. After pursuing the matter for one year, the researcher could receive 60% response from Indian universities. At the same time researcher obtained maximum information from DLI-2 projects by visiting their websites.

6.2 Analysis

After analyzing the Indian responses, it was found that out of 132 respondents only 20 respondents affirmatively said that they have undertaken some type of digitization activity. Though some of them had mentioned that they have undertaken digitization activity, it was observed that they have not undertaken the digitization activity but are carrying out computerization modernization of libraries. Out of 20, twelve respondents clearly mentioned that their digitization activity is at planning stage. i.e. some have prepared the proposals and sent for approval etc. It was found that only eight Indian universities are actually engaged in the work of digitization and digital libraries. i.e. they have acquired the computers, scanners and other infrastructure required to carry out the project. The list of these organization is given in Table 6.1. DLI-2 projects list is given in appendix-A.

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>Name of University</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>University of Hyderabad (UOH)</td>
</tr>
<tr>
<td>2</td>
<td>Gujarat Ayurveda University, Bhavnagar (GAU)</td>
</tr>
<tr>
<td>3</td>
<td>University of Pune (UOP)</td>
</tr>
<tr>
<td>4</td>
<td>Anna University, Chennai (AUC)</td>
</tr>
</tbody>
</table>

Table 6.1 Indian Universities having digitization projects
Project title
After analyzing the answers it was found that out of eight Indian respondents only three have the specific title for their digitization project. Whereas all 28 DLI-2 projects have specific project title. The project title itself indicates the nature of work, type of research work is being carried out by that center. The details of the project titles are given in Table 6.2.

### 6.2 Project title

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>University</th>
<th>Project Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UOA</td>
<td>High-Performance Digital Library Classification Systems: From Information Retrieval to Knowledge Management</td>
</tr>
<tr>
<td>2</td>
<td>UCB</td>
<td>Re-inventing Scholarly Information Dissemination and Use</td>
</tr>
<tr>
<td>3</td>
<td>UCD</td>
<td>A Multimedia Digital Library of Folk Literature</td>
</tr>
<tr>
<td>4</td>
<td>UCLA</td>
<td>Cuneiform Digital Library Initiative</td>
</tr>
<tr>
<td>5</td>
<td>UCSB</td>
<td>Alexandria Digital Earth Prototype</td>
</tr>
<tr>
<td>6</td>
<td>UCM</td>
<td>Informedia-II: Auto-Summarization and Visualization Over Multiple Video Documents and Libraries</td>
</tr>
<tr>
<td>7</td>
<td>CU</td>
<td>A Patient Care Digital Library: Personalized Search and Summarization over Multimedia Information</td>
</tr>
<tr>
<td>8</td>
<td>COR</td>
<td>Project Prism at Cornell University: Information Integrity in Digital Libraries</td>
</tr>
<tr>
<td>9</td>
<td>EC</td>
<td>Digital Analysis and Recognition of Whale Images on a Network (DARWIN)</td>
</tr>
<tr>
<td>10</td>
<td>HU</td>
<td>Operational Social Science Digital Data Library</td>
</tr>
<tr>
<td>11</td>
<td>UHM</td>
<td>Shuhai Wenyuan Classical Chinese Digital Database and Interactive Internet Worktable</td>
</tr>
<tr>
<td>12</td>
<td>UIC</td>
<td>Digital Library for Human Movement</td>
</tr>
<tr>
<td>13</td>
<td>IUIB</td>
<td>A Distributed Information Filtering System for Digital Libraries</td>
</tr>
<tr>
<td>14</td>
<td>IU</td>
<td>Creating the Digital Music Library</td>
</tr>
</tbody>
</table>
From the above table it is clear that efforts are being carried out to find out solution with respect to digitization of folk literature, patient care information, whole images on a network, social science digital data library, digital library for human movement, creating the digital music library, indexing handwritten manuscripts, creating digital library for humanities and automatic reference librarian for the world wide web which indicates that multi-disciplinary subject are undertaken by these projects.

**Objectives of the project**

Objectives clearly indicate what type of work is carried out and what is going to be achieved by undertaking that projects. The objectives of one of the Indian Universities are to digitize old and rare books, to develop a Marathi metadata to describe Marathi documents, to provide wide access to scanned books and to provide technological help to other organizations in scanning and planning digitization project, where as the objectives of the DLI-2 projects are depicted in Table 6.3.
<table>
<thead>
<tr>
<th>University</th>
<th>Objectives of projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>UOA</td>
<td>to develop an architecture and the associated techniques needed to automatically generate classification systems from large domain-specific textual collections and to unify them with manually created classification systems to assist in effective digital library retrieval and analysis</td>
</tr>
<tr>
<td>UCB</td>
<td>to provide a testbed for computer science research in image analysis, digital documents, and information retrieval</td>
</tr>
<tr>
<td>UCD</td>
<td>to digitize the folk literature of Jews</td>
</tr>
<tr>
<td>UCLA</td>
<td>to provide access to the text catalog of the Cuneiform Digital Library which contains information about nearly 54,000 cuneiform documents from the 4th and 3rd millennia B.C.</td>
</tr>
<tr>
<td>UCSB</td>
<td>to build a distributed digital library for materials that are referenced in geographic terms, such as by the names of communities or the types of geological features found in the material</td>
</tr>
<tr>
<td>CMU</td>
<td>To provide full content search and retrieval of current and past TV and Radio news and documentary broadcast</td>
</tr>
<tr>
<td>CU</td>
<td>to create a fully-integrated, highly selective, interactive online resource for educational materials in this field, and to evaluate the ongoing value and economic viability of providing these services on a subscription-based, cost-recovery model to educational institutions, libraries, and students</td>
</tr>
<tr>
<td>COR</td>
<td>to investigate and develop the policies and mechanisms needed for information integrity in digital libraries</td>
</tr>
<tr>
<td>EC</td>
<td>allows marine scientists to maintain information for the study of various behavioral and ecological patterns of bottlenose dolphins</td>
</tr>
<tr>
<td>HU</td>
<td>To create a user-friendly, large-scale network of social science data</td>
</tr>
<tr>
<td>UHM</td>
<td></td>
</tr>
<tr>
<td>UIC</td>
<td></td>
</tr>
<tr>
<td>IUIB</td>
<td>To deal with heterogeneous information sources that may change over time handling shifting requirements and interest of users with minimal user involvement creating a robust agent architecture for reactive, proactive, and collaborative filtering supporting agent collaboration based on natural or artificial economic frameworks for multiagent tasks</td>
</tr>
</tbody>
</table>
From the above table it is clear that these objectives specifically mentions the nature of work of their projects. UCD's objectives is to digitize the folk literature of jews, while CMU's objective is to provide full content search and retrieval of current and past TV and Radio news and documentary broadcast. UMA had a objective to develop techniques for indexing handwritten historical manuscripts.
Duration of the projects:

As digitization is discussed at many platform in India in recent years, it is observed that most of the projects have recent origin. Only one project has noted that their project had started in 1984. Remaining seven projects are started around 2000 and 2002. As compared to Indian situation most of the DLI-2 projects have started in 1999. The details are given in table 6.4

Table 6.4 Duration of project

<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>University</th>
<th>Start Date</th>
<th>Closing date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UOA</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>2</td>
<td>UCB</td>
<td>1999</td>
<td>2004</td>
</tr>
<tr>
<td>3</td>
<td>UCD</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>4</td>
<td>UCLA</td>
<td>2000</td>
<td>2003</td>
</tr>
<tr>
<td>5</td>
<td>UCSB</td>
<td>1999</td>
<td>2004</td>
</tr>
<tr>
<td>6</td>
<td>UCM</td>
<td>1999</td>
<td>2003</td>
</tr>
<tr>
<td>7</td>
<td>CU</td>
<td>1999</td>
<td>2004</td>
</tr>
<tr>
<td>8</td>
<td>COR</td>
<td>1999</td>
<td>2003</td>
</tr>
<tr>
<td>9</td>
<td>EC</td>
<td>2000</td>
<td>2002</td>
</tr>
<tr>
<td>10</td>
<td>HU</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>11</td>
<td>UHM</td>
<td>2000</td>
<td>2003</td>
</tr>
<tr>
<td>12</td>
<td>UIC</td>
<td>2000</td>
<td>2003</td>
</tr>
<tr>
<td>13</td>
<td>IUIB</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>14</td>
<td>IU</td>
<td>2000</td>
<td>2005</td>
</tr>
<tr>
<td>15</td>
<td>JHU</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>16</td>
<td>UOK</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>17</td>
<td>UMA</td>
<td>2000</td>
<td>2003</td>
</tr>
<tr>
<td>18</td>
<td>MSU</td>
<td>1999</td>
<td>2004</td>
</tr>
<tr>
<td>19</td>
<td>OHSU</td>
<td>1999</td>
<td>2001</td>
</tr>
<tr>
<td>20</td>
<td>OUP</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>21</td>
<td>UOSC</td>
<td>1999</td>
<td>2003</td>
</tr>
<tr>
<td>22</td>
<td>SU</td>
<td>1999</td>
<td>2004</td>
</tr>
<tr>
<td>23</td>
<td>UTA</td>
<td>1999</td>
<td>2002</td>
</tr>
<tr>
<td>24</td>
<td>TU</td>
<td>1999</td>
<td>2004</td>
</tr>
<tr>
<td>25</td>
<td>UOW</td>
<td>1999</td>
<td>2001</td>
</tr>
</tbody>
</table>
It is observed from table 6.4 that maximum duration of the project is six years i.e. mentioned by MSU, as its objective is to create a significant, carefully organized on-line repository of spoken word collections. Where as minimum duration of the project is two years of UOW and OHSU Universities.

**Funds for the project:**
Long-term access to digital collections depends upon careful life-cycle management. Funding plays an important role in this life-cycle. Funding is required not only for the creation of digital materials but also for metadata, storage capacity, preservation tools, navigational tools and user supports. One has to find out also how long the project is going to be supported by funding agency. What plans exist to make it self-sustainable etc.

It is observed that most of the Indian project had not mentioned the budget provision. Out of 8 respondents only one has mentioned that Govt. of India made a provision of Rs. 3 lacs per year. On the other hand all other had received funds for these projects. ( which are mentioned in table( 6.6 ) by funding agencies such as NSF,
From the above table it is clear that the lowest fund received by E.C i.e. $32,870 while highest budget received for digitization by CMU i.e. $5,400,000 & UCSB.

**Purpose of digitization**

Digitization presents the opportunity to enhance access to information throughout the world. The purpose for the creation of digital products varies from institution to institution. The most commonly purpose for digital products was educational interest in Indian context. Out of eight respondents six replied that their purpose of digitization was for education purpose while four said that they feel digitization can be used as a preservation tool for their old and fragile material. At the same time four respondent feel that digitization can provide more access.
As compare to Indian situation most of the DLI-2 projects had been undertaken from educational point of view. Out of 28 project 17 universities are creating digital information for educational interest while 11 have undertaken digitization as a preservation tool to preserve their old materials, 14 organization opted for more access, and information about 3 projects is not available.

**Digitization of materials**
Digitization makes it possible to capture vide array of information types and present in the most appropriate format. It is found that all eight Indian universities are engaged in converting text into digital format 3 have undertaken the digitization of photographs while four (50%) organizations are busy with manuscripts work. Digitization of map is undertaken at one university while microforms, audio – video material is neglected in all these universities.

Most of the DLI-2 projects are initiated taking into consideration the specific need of user community. Their objective was to cover all human endeavors. Due to this all these 28 projects have undertaken the digitization of diverse field. Which include text, manuscripts, microfilms, music, audio-video cassettes, maps, etc. out 28 projects 25 are devoted converting text into digital form while 17 are working with digitization of photographs. Manuscripts seem too little difficult as only 4 universities are Dealing with it. Most of the projects are dealing with text, photographs and manuscripts which is diagrammatically represented in figure 6.1
Nature of content

There is a variation about nature of context of the material content can be generally classified as historical information, scientific information, technological information, culture heritage, artistic information etc. With respect to Indian responses, it is clear that out of eight universities six universities had chose the original material from the historical point of view. Second priority goes to scientific and technological information, which indicates 50% institutes have undertaken digitization to convert scientific and technological information into digital form.

As compared to Indian response most of the DLI-2's efforts were selecting all type of contents, which are mentioned above. The information reveals that out of 25 projects 19 projects clearly gave priority for scientific information and second priority is for historical information followed by technological and cultural information. While only 7 projects had devoted their efforts to concentrate on converting artistic materials into digital form.

Figure 6.2 Type of digitized material

Nature of content

There is a variation about nature of context of the material content can be generally classified as historical information, scientific information, technological information, culture heritage, artistic information etc. With respect to Indian responses, it is clear that out of eight universities six universities had chose the original material from the historical point of view. Second priority goes to scientific and technological information, which indicates 50% institutes have undertaken digitization to convert scientific and technological information into digital form.

As compared to Indian response most of the DLI-2's efforts were selecting all type of contents, which are mentioned above. The information reveals that out of 25 projects 19 projects clearly gave priority for scientific information and second priority is for historical information followed by technological and cultural information. While only 7 projects had devoted their efforts to concentrate on converting artistic materials into digital form.
The long-term access to digital materials depend upon the nature and type of storage media used. It is observed that there is tremendous growth and change with respect to hardware and software. The life cycle of digital media depends upon the refreshing and migration of data. From the data received it is evident that all eight Indian universities are storing their information on hard disk. At the same time its backup is taken on CD-ROMs. Out of eight organization two are using magnetic tape for storing digital information.

With respect to DLI-2 it is observed that information about hardware and software is not available on their websites. But the researcher tried to collect this type of information by visiting the specific university's home page. It is observed that most of the projects are using Pentium 4 for their digitization project. University of California Berkeley is carrying out 40 different digital library projects. This university is using hardware provided by Sun Micro system. The specification of this system is as – HP (D4366N) xu 6 / 200 Pentium pro with 4 GB SCSI / 160 MB – BRAIN DEAD, they are using HP Document scanner. Following software are used for this project

- Allegro CL
- Net worker and Autochangeer UNIX V. 5.2.2.
- Office 2000 PRO
- Paint shop PRO
- Photo shop
- Visual café Java
- Corel photo CD collection
- Secure CRT
- Salaries 7 server (450) and Doc.
- Jbuilder PRO
- Frame maker (Unix, Sun and HP only)

As compared to foreign organization all 8 universities from India are using Pentium 4 and most of them are using flatbed scanners only at one place Minolta Book Scanner is used.

File formats

The future migration of data depends upon the file format. HTML or postscript files can be converted to Adobe's portable Document format (pdf) Typical standard formats is ASCII for text, TIFF for image, jpeg or mpeg are audio and video etc. It is observed that out of eight Indian universities only two replied to this question. The information reveals that these two universities are using html, pdf, gif, jpeg and tiff format.

DLI-2 projects are using multiple file formats. The most common file formats are html, pdf, gif and tiff. Those projects, which are dealing with audio and video materials, are using jpeg and mpeg file formats. Available data is represented in the following table 6.5

<table>
<thead>
<tr>
<th>File type</th>
<th>No. Of universities</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTML</td>
<td>25</td>
</tr>
<tr>
<td>Sgml</td>
<td>20</td>
</tr>
<tr>
<td>Pdf</td>
<td>18</td>
</tr>
<tr>
<td>Ps</td>
<td>15</td>
</tr>
<tr>
<td>Jpeg</td>
<td>12</td>
</tr>
<tr>
<td>Gif</td>
<td>24</td>
</tr>
<tr>
<td>Tiff</td>
<td>23</td>
</tr>
</tbody>
</table>

Table no. 6.5 Use of different files

260
Navigational tool

The retrieval of information on the computer has become very easy with the help of navigational (search) tools. The availability of free access to some Internet resources and retrieval of remotely located information has become possible through navigational tools only. It is found that the projects included in DLI-2 have developed their own software to retrieve their information.

With respect to Indian respondents it is found that only two universities out of eight have developed or using readymade software for information retrieval. VTLS is one of the most popularly used software university of Hyderabad for information retrieval.

Project personnel

Personnel’s are required from the planning stage of the project to the every day maintenance of file creation and back-up. Most of the DLI-2 projects have full fledged project personnels including project Director, co-ordinator, programmer, Assistant-programmer and Assistants for scanning and subject experts for editing of files etc. It is also observed that most of
the Directors and co-coordinators of these projects are Heads of the computer science Department.

Most of the Indian project (7 out of 8) is managed by the existing staff. Only at one university a separate staff is appointed to look after digitization activity.

Researcher would like to know the opinion of authority about, the digitization as long-term solution.

**Digitization as long-term solution**

Many are of the opinion that digitization can be a long term solution for preservation of library materials. It is observed that out of eight Indian respondent only four feels that it would be a long-term solution since there is a change in hardware and software set-up. Technological obsolescence is major threat to the digital information. Many experts are suggesting refreshing, migration as a preventive measure to ensure the preservation of digital information for longer time. But it is observed that there is some sort of loss when one file format is converted to another format for example if ppf file content is converted to MS-word, the original document is mixed with many junk characters and figure, tables and pictures are converted to ASCII characters which are of no use to users.

**Problems neutered**

Most of the Indian universities identified following problems----

- Scanning old books are very difficult and expensive.
- Retrieval of information from Devanagari manuscript is extremely difficult.
- Scanning of over-sized books, maps and photographs is difficult.
Scanning of Theses: there was a problem in scanning old theses, because of its old texture problem. It leaves blank spot on the scanned pages. While using ADF scanner bounded theses have to be rebound even there are problems where the page contains text, image, mathematical and chemical equations and formulae.

- Lack of technical expertise
- Lacking of standardization and proper guidelines.
- Inadequate budget provision
- Failure of hardware / software
- It is time consuming- tasks long time from scanning to transmit (chain)
- Maintenance of foreign hardware (scanner) is difficult since local personnel's are not trained to repair these machines.
- Copyright for books and journals
- Setting of different size book on scanner is time taking
- Ending files taking too much time
- Eye straining while scanning documents continuously.

6.3 Digital Library Models

The retrieval of information on the computer has become very popular in recent years. The availability of free access to some Internet resources and the rapid expansion of Internet, particularly in the recent years, have made the retrieval of remotely located information a popular exercise. The availability of free access to systems like World-Wide Web (WWW) even allows users to incorporate their own relatively easy-to-implement hypermedia systems into WWW to become part of an integrated Internet resource system. Several different Internet resource systems have been linked with each other in reasonable, though not ideal, ways. The realization of a practical digital library system is becoming a reality. An
example of subsystem implementation is WATERS, (Wide Area Technical Report Service) jointly developed by ODU, UVA, SUNY Buffalo, and VPI. In that system WAIS is combined with WWW allowing anybody with access to Internet to search a distributed database of computer science technical reports with subject keywords, authors or university names, and issue dates among others. The user does not have to worry about where a particular technical report is stored and retrieval is optimal in the sense that technical reports are stored locally under the control of the authors and sent directly by WWW to the requester once WAIS has identified the location.

To a large extent, Internet collects and organizes information around administrative units. Many organizations that have benefited from the Internet also want to share their knowledge with others. They put the information into some repositories and try to make the information generally available. Because the administration of the Internet's networks is decentralized, no single organization has a clear idea about which resources are available in the Internet to its end-user and how to guide the end-user to the information of interest. The large amount of the resources and the lack of guidance for their access has posed great challenges for users to find, acquire, organize, retrieve, and use the information. Besides other technical aspects, a central problem for designers of document and information retrieval is often related to the contents of the documents rather than their organization. The contents or semantics of documents are not well represented by only surface features such as individual words taken from abstracts, titles, or even when taken from the entire documents. The problem is complicated further when the type of information is image, audio, video, or other media, where even the surface features are not generally available.

The existing systems (Archie, Gopher, WAIS and WWW) [4] are still largely inadequate in connecting the information needs of the end-user to
the vast Internet resources, given the diversity of user classes and resource representations, the constraints on autonomous administration of the resources and on the heterogeneity of information systems, applications, and user interfaces. Internet information discovery tools need non-trivial approaches to map the massive administrative-centered resources into a kind of conceptual information space that will closely reflect the user's information needs. This clearly indicates a need for ideal digital library.

Before building a digital library one has to formulate a clear conception of what is to be achieved through a proposed digital library. This is called the ideology of organization. The ideology can be formulated in terms of the collections purpose, the objectives it is intended to achieve, and its principles, the directives that will guide decisions on what should be included and what should be excluded. Whenever one builds a digital library collection, one should formulate its purpose and state it clearly as an introduction to the collection. One should also make it plain to users what principles have been adopted to govern what is intended. It is also equally important to describe how the collection is organized. In the day-to-day life everyone comes across several models consciously or unconsciously. Some take them as real and some try to ignore, and others use them with advantages. Models are increasingly gaining importance these days. They are considered to be fascinating and are used widely. Most of the projects, operations etc. construct their own models that relates to their courses of action. Each model has its own unique features, characteristics and characteristics. And usage. It is equally important to understand what model is and why how it is constructed and used. An effort is made here to examine and understand the meaning and concept of model, the procedure involved in designing model, the ingredients and the types etc.
6.31 Concept and meaning of a model

A model is a representation of a particular reality or phenomenon in any field of activity or process. The reality can be represented in various ways by way of modeling. There are quite few definitions for a model. However, most of them underline the basic concept that it is a meaningful representation of a reality.

Oxford English Dictionary [1] defines models as ‘a simplified description of a system, process etc. put forward as a basis for theoretical or empirical understanding; a conceptual or mental representation of something’. A model is thus an abstract form of a particular real phenomenon.

A model is a purposeful representation of reality defines Douglas D. Mooney and Randall Swift [2] They make it clear that model is prepared for a specific purpose to achieve and represent a particular reality.

A model is a description of the interdependencies of a phenomenon in terms of mathematics, pictures, computer programming language or some similar descriptive language, together with a theory of the dynamics of the subsystem. It is a description, in abstract terms of, interdependencies among some of the individual's functions, and activities in an entire system or in some part of the same system.

There are different types of models that are in use. Each model is unique in nature and has its ingredients. Further, it has purpose to serve. Several models are being developed and used. They may be categorized, depending upon the objectives, purpose, form and other features. They are mainly divided into measurement models, decision-support models and theoretical models, which are used for measuring, decision-making and theory building.
Nürnberg [3] considered three broad classes of library elements: i.e. data, metadata, and processes. Data are library materials. Metadata are information about the library and its materials. Processes are active functions performed over library elements. It is easy to find examples of physical library data that are translated into digital form routinely. For example, books, journals, and movies are all examples of physical library data that are scanned, digitized, or otherwise translated and put online.

Digital library architectures include many different types of resources that come into contact with each other at different times and for different reasons. The interactions between elements may vary by resource, by user, and may even have temporal aspects. Consequently, these architectures are a particularly challenging concept to represent graphically. In this respect, Alan Pagliere [4] suggested a model of representation whereby two elements that have a direct relationship with each other "touch" each other, which is shown in figure 6.1. The model worked very effectively, allowing the group to rapidly talk through a number of scenarios and modify the drawing in an iterative process. This process ultimately resulted in the graphic, which accommodated the various collection/access scenarios proceeded to describe. The triangular or wedge shaped objects shown in the figure are understood to have contact with elements along their sides and not their points. So, for example, Index has contact with Class Middleware and Repository, but not Object Name Resolution.
Conceptually, a digital library system may be thought of as mediating certain kinds of interactions among people and computing systems. Figure 6.2 shows some relationships and interactions among several parts of the digital library and several people and systems external to the library. To help clarify the interactions occurring in these relationships, the computing resources in this figure have been partitioned into server resources and client resources. This allows the classification of computer-supported relationships into human/human, human/client, human/server, and client/server classes.
The real relationships are often more complicated than shown. For example, publishing in the digital library is not strictly a relationship between publisher, librarian, and the digital library server. Patron needs, budgetary constraints, limitations of library computing resources, and a number of other factors may be involved.

6.321 The University of Michigan Digital Library Model

The University of Michigan Digital Library (UMDL)'s [5] prime consideration in proposing the new digital information architecture was the ease and flexibility of its implementation. In particular, they are concerned with the heterogeneity of users' hardware, the heterogeneities in the underlying networks (e.g., differences in speeds), and the autonomous nature of individual repositories. These considerations are reflected in the architecture as shown below. The implementation issues are discussed in terms of its three layers which is shown in the figure 6.5
Figure 6.5 Interoperability architecture for Digital Information Repository
a variety of commercially developed interfaces to meet these needs. Some of the major issues in developing such interfaces are:

a) **The level of transparency** offered to the user in terms of the locality of resources, the cost of access, the boundaries of search domain, etc.

b) **Accuracy of the retrieved information**: Often users want a quick response to a query, even if approximate, rather than wait a long time for the most accurate answer. For example, an economist studying the economic trends over the last 100 years does not care whether the data is accurate up to yesterday or last week. The interface should offer an option for the user to specify the required accuracy. The system can in turn use this information to choose appropriate copy.

c) **Complexity of the interface primitives**: One of the key design and implementation decisions at this layer is the nature of primitives. One approach (e.g., RISC-like) is to offer simple and efficient primitives to the user interface so that customized UIs can be built using them. In this case, cost of building UIs will be high.

d) **Representational issues**: Since this layer has to handle all types of data including textual, pictorial, and audio (data, voice, and video in the network terminology), a decision has to be made as to the role of this layer in handling the data. If a uniform access approach is adopted, then it could handle all the data just as one kind: digital. In this case, the user interface will be responsible for recognizing the differences in the data types and handling them accordingly. Some of the issues that pose implementation difficulties are the management of buffers, management of computation and communication resources, and dealing with any synchronization requirements. For example, if the information retrieved from the underlying system consists of independent audio and video sources, but to be presented in a synchronized manner to the user, there may be additional
efforts necessary at the user interface layer. It is also possible to transfer this responsibility to the UIS.

**Interoperability Layer**

This layer interfaces with a variety of underlying resources and offers a uniform interface to the user interface layer. Some of the key implementation issues to be dealt with at this layer is:

- **a) Tools:** In order to implement a variety of search/retrieve/combine operations over a large domain and heterogeneous resources, it is needed to integrate several existing commercial products. UMDL certainly foresees the need for hypermedia tools, collaboration tools, large-space search techniques, etc. Selecting the tools as well as integrating them to achieve the desired task may be quite a challenge.

- **b) Interface routines:** The IL protocol has to deal with a slew of resource clusters. Once again, depending on the uniformity (or non-uniformity) offered by the interoperability layer, the routines could become quite complex. Problems of heterogeneity, problems of size (large set of repositories to deal with), and the complexity of services offered to the user interface call for innovative design and implementation. Especially, one should be concerned about the end-to-end delay expected by the users. The implementation choices will be greatly dictated by the performance requirements of the end-user.

- **c) Dealing with autonomous repositories:** Since each repository is autonomous, there is no standard format for the data that it stores or the primitives that it offers to the rest of the system. This creates problems in the implementation of standard RRP's.
Resource Repository Layer

This is the layer in which UMDL have least control on-since retaining individual repository autonomy was their objective. However, to facilitate its integration with the rest of the system, UMDL proposed a flexible interfacing system. The interface would be developed by the repository maintainer. However, the digital library administrators would provide guidelines as to the minimal services expected from each repository. Whether the semantics (and syntax) of these minimal services are to be standardized and set by the digital library system administrators is not clear at this point. Such standardization at least on the minimal subset would ease the development of the interoperability layer. In addition, it will guarantee some services from each repository. The issue of charging a fee for access is the discretion of individual repositories. UMDL do not intend to look into the issues of privacy and charging by the local sites.

Figure 6.3 illustrates, at a very high level, the UMDL, linking several users through their User-Interface (UI) agents to collections through Collection-Interface (CI) agents. In a simplistic system, the UI agents and CI agents could be networked together, allowing UI agents to query collections directly, either sequentially or in parallel. There are many problems with this simple solution, such as the duplication of effort in having UI agents determine the subset of CI agents needed to service a particular request, or the complexities of terminating the search once a CI agent has successfully answered a user query.
6.322 National Digital Library Model (NDL)

Library of Congress [6] had developed an informal model shown in figure 6.4 for access to the NDL. This diagram presents a general model through specific examples that are implemented in some form today. The model allows the user to identify relevant resources by searching indexes or by browsing documents. Automatic links from indexes or documents to an individual digital item are implemented by use of its unique name. A name (or rather, in today's implementation, the corresponding URL) can also be used directly by the user to retrieve the item without performing search historical collections.
6.323 The Illinois Digital Library project

The Illinois Digital Library project [7] is constructing a large-scale digital library for engineering documents and databases. This project consists of two inter-related parts. The first is building a testbed of materials obtained from professional and commercial publishers, with software that will be used by an engineering community of thousands of users. The second is performing research in technology and in sociology to understand how to scale the testbed model to the National Information Infrastructure.

The digital library itself will be centered on a collection of engineering journals and magazines, obtained through collaboration with a range of major professional and commercial publishers. The intention is to attract a
broad range of usage from a broad range of users. All documents will be structured and complete, that is, encoded in SGML and containing all pictorial material. The documents will include general engineering magazines (e.g. computer science from IEEE), specific engineering journals (e.g. aeronautical engineering from AIAA), and specific scientific journals (e.g. physics from APS). Finally, articles from commercial engineering publishers (e.g. Wiley & Sons) will be collected for use in their economics trials.

In addition to the document collections, a number of databases will be gathered into the digital library and cross-linked to the documents where possible. These include significant databases generated by other NSF-funded projects, e.g. the BIMA Grand Challenge database in radio astronomy and the WCS National Collaborator database in molecular biology. Associated GIS satellite image databases include the NASA-funded data supported by the NCSA HDF project. These projects are local to the University of Illinois and supervised by collaborators on the digital library project.

The testbed software will go through two primary phases within the proposal period of four years. The goal of version 1 is to leverage off their substantial existing resources to build a functional digital library with a large collection used by a substantial user population.

6.324 Columbia University (Janus) Digital Library Model

Janus [8] is a prototype digital library of Columbia University, which was established in 1990. Janus is providing coherent access to multiple document types, including text, images, video and combinations. This initiative includes research in user interfaces, search and retrieval of text and images, representation of both text and images to facilitate search and delivery, advanced multimedia networking protocols, and research on
representation and reasoning for different models of intellectual property rights. Parallel to each of these research efforts, research on evaluation with users is provided to have early feedback to each component, shaping the research design, which ultimately quantifies successes in each field. An overview of Janus system is shown below.

Figure: 6.5 Janus model

The Janus user interface integrates text, image, and other media for both query formulation and response, including the ability to automatically generate summaries of the retrieved documents that coordinate natural language and graphics. The testbed includes text, images, and video, as well as documents that integrate media, such as annotated legal documents; scientific documents containing photographs or diagrams, and humanities documents containing artwork. The aim was to develop integrated search and retrieval of the multimedia testbed, providing
integrated indexing of text and images to allow retrieval of a textual document, an image, or both in response to a single query. In order to support effective search and retrieval as well as summary generation, this system provides information about the semantics of document segments in both texts and images.

Once the user has selected images and documents of interest, the intellectual property rights inference system will determine the rights of the current user and the different options available for obtaining copies of the document (e.g., free or pay-per-use, where use can include different charges for printing or online browsing).

6.325 Digital Library Model by Srihari

Srihari and others [9] envisioned DL as a server connected to a computer network of users where the server has the ability to respond to user requests by retrieving relevant information from document data stored on an electronic storage medium. The specific research topics are determined by the architecture of the DL, which is shown in Figure 6.6. It consists of an off-line process wherein document data is captured and prepared for storage, and an on-line process where queries are received and processed for information retrieval. There are three major computational processes shared between the off-line and on-line processes of the DL viz. document data capture, data integration/indexing and information retrieval.
Digital libraries can be created as a blank form into which virtually unlimited information can be loaded. Another question arises, what should this digital library look like? What are the criteria to judge whether or not a particular model of the digital library is successful implementation of the concept. Ken Varnum [Varnum] proposed three general requirements that must be met if a digital library is to be considered a success. The implementation must be scalable, useful and most important is useable. Any digital library design should be able to expand and handle many more resources, it should provide access to them and it should handle many users. The most popular WWW search engine, Lycos which came online in May 1994 as part of the Carnegie Mellon University’s Computer Science Department server, has grown from about 30 accesses/week to over 120,000 accesses/week by November 1994. The idea of useful is subjective and there are no concrete measures by which a system’s usefulness can be judged effectively unless it is up and running. The University of Michigan (UMDL) proposal recognizes the failure of traditional means of measuring usefulness of material culled from a
database, and by extension, of the database itself. The only way to judge the usefulness of the digital library is to allow it to compete with other digital libraries, all of which provide access to the same information resources and to see which one is used more.

Usability is less simple to define than scalability. In the latter’s case, it is clear if a system is scaleable--if it can handle increased loads on both ends (user and provider), then it is scaleable. If the users can find easily information on a topic or if the system allows users to find information on a topic of interest to them, then the system is most usable. A useable digital library will have information in a variety of formats and for a variety of purposes form cursory to introductory to thorough and intensive. The access interface should be sufficiently easy; otherwise no good results will be achieved. If users cannot instruct the computer what he/she wants, then the system is not user friendly. The interface must be intuitive and must be flexible enough so that advanced users can access more advanced functions. A WWW browsing tool like Netscape and Lynx might be an appropriate model for the interface used at a lower level of library.

6.326 Digital Library model with rich semantic structure

A digital library model with rich semantic structure was proposed in the area of Alcohol and other Drugs (AOD) substance Abuse and Prevention, assembling and interlinked information from several government and private agencies [10]. This model is designed to specify an information structure and an integrated set of functions links to where work on each function has been or is being done. The objectives of this modal were.

- To develop a blueprint for an enriched digital library.
- to implement and test a basic version of this blueprint, incorporating existing tools wherever possible, with emphasis on simplicity and basic functionality in a toolbox that can be deployed easily.
to make available the functionality of the system with a small sample collection, including the alcohol and other Drug Thesaurus.

to test the usability of the system and get a glimpse at the ways in which users Internet with the system and use it for their work.

To achieve these objectives, the project built a total environment making rich semantic structure available to the user. The main goal was to collect simply what is available on the web. This information can be collected from various federal agencies. Central to this system design was support for the users' work and for information sharing and collaboration. Users were able to contribute the public information space in which they could store queries, documents, annotations, links, subject-indexing information etc.

The information structure envisioned covered several object type (such as concepts, documents, projects, persons, organization and queries) and link types (such as deals with defines, produces, includes cites). The functions or tools are divided into search tools, especially a tool for browsing concepts hierarchies, a search form for searching the system and external database, thesaurus support for searching, and glossary tools; document creation and editing tools, supporting collaborative authoring and annotation and link management; and document processing and lynk generation tools, including a document segmentation tool to structure plain text, a citation index tool, a summarization tool and a document categorization and indexing tool which can be visualized form figure 6.7

6.327 Classical Digital library model (CDLM)

Classical Digital Library Model (CDLM), which was developed, by Steven L.Mcgall and others [11] is derived from traditional practices of library and information science professionals. These practices include the expert
critical selection and organization of the networked information resources for local populations of users and the integration of advanced information tools such as databases and the Internet. Into these networked collections. The CDLM examines what professional services can be delivered to the desktop computers of users. The strength of the CDLM is that it enables LIS professionals to extend their practices to the digital library environment where networked delivery of digital content does not require library users to be physical present in a library. The CDLM is a service oriented model developed so that the best digital resources appropriate for users and the needs arising from their usual processes, can be integrated by LIS professionals who have appropriate local intelligence. The scope of the CDLM encompasses the digital libraries' development and management from needs assessment, information selection and organization to access and use.

Under the CDLM, users do not have to browse, navigate or search large and diverse information space without the assistance of LIS professionals who are able to pre-position selected quality resources and organize them with their need in mind. The digital library acquires content, which is suited to the needs of the users as determined by local library professionals. The main goal of developing a CDLM was to provide a digital library environment within which a typical user can autonomously browse for documents relevant to specific information needs. There are three principle aspects of this i.e. wayfinding, collection development and organization of resources. Wayfinding is a concern from the physical library model. The premise of library wayfinding is that users who are able to help themselves are much more likely to be return visitors. How should a digital library be designed so that users are able to help themselves with minimal guidance and thus revisit on a regular basis? Elements of wayfinding, collection development and organization of information play a shared role in this model.
6.328 Miami University digital library model

The Digital Information Services of Miami University [12] Cluster provides a framework for proposing, creating, promoting, publishing, and maintaining Digital Collections.

Within the Miami University Libraries, a Digital Collection is defined as a project, which has these characteristics:

- Organizes materials in digital format which support reference, instruction and research at the University
- Contains or will eventually grow to contain a number of items which are substantial in size and importance to the Miami University Community
- Includes objects which are searchable and/or can be included in a searchable database of other digital collections
- Will be made available through web-accessible pages which conform to the standards of other Digital Collections at Miami University

6.329 Proposed Digital Library Model

After taking into consideration the various models studied in detail, the researcher has come out with an ideal digital library model for university libraries.

The components of this model are already discussed while studying other digital library models. Computer terminals are available to access the information. The library staff has to perform two types of jobs. First if they do not have scanners and other infrastructure to convert the existing...
materials, they can download the information from the Internet or CD-ROMs. This digital data can be acquired as per the needs of the users. This digital information can be made searchable using some type of navigational software. The researcher recommends the Greenstone digital library software because it is free of charge and easy to install and use. The another task the library staff has to perform is of selecting the valuable and important documents for digital conversion. The detailed procedure about this is fully described in chapter three. Since the scope the present study is limited to the rare materials, effort has been made to convert few documents using Minolta PS 7000 scanner. There are many standard website which provides reliable information online. This information is collected and made accessible using GSDL. The researcher have used flatbed scanner and minolta ps 7000 to convert some rare books. Around 20 manuscripts and 20 rare books have been converted into digital form. Each page is saved as a separate file in different file formats. The researcher had downloaded many articles and projects reports from Internet. Hence GSDL is tested for both type of material i.e. born digital and converted material. It is found that GSDL converts pdf, ps, ms-word etc. into html format. The search is very powerful but users have to gain some knowledge of PERL language to make effective use of GSDL. The output of this experience is appended in appendix-D.

After taking into account all these facts about GSDL, the researcher is of the opinion that any librarian having some basic knowledge of computer applications can create the digital library for their users. If this digital library is made available through their websites, other users looking information on same topic will be benefited. Librarians can acquire or convert digital information taking into considerations the need of their users. And DL software have made it possible to manage and organize the large amount of digital information. The findings of the present research are given in the nest chapter.
References: