
UNIT 7 DIGITAL LIBRARIES

Structure

- 7.0 Objectives
- 7.1 Introduction
- 7.2 Digital Library: Concept and Definition
- 7.3 Characteristics of Digital Libraries
- 7.4 Development of Digital Libraries
- 7.5 Digital Libraries and their Uses
- 7.6 Major Issues/Challenges
- 7.7 Summary
- 7.8 Answers to Self Check Exercises
- 7.9 Keywords
- 7.10 References and Further Reading

7.0 OBJECTIVES

In recent years the concept of digital libraries is gaining lot of importance amongst library and information professionals and users as well. Digital libraries have been emerged which will facilitate storage processing and accessibility of information much more efficiently than conventional libraries.

After reading this Unit you will be able to:

- understand the concept and development of digital libraries ;
- comprehend the characteristics of digital libraries ;
- understand techniques of building digital libraries; and
- know the issues/challenges before the digital libraries.

7.1 INTRODUCTION

The term digital library has a variety of meanings ranging from a digital collection of material that one might find in a traditional library to the collection of all digital information along with the services that make that information useful to all possible users. Digital libraries have evolved as a result of fast technological development in order to cater to the needs of individuals with varying interests in various fields. Although the term digital library has gained popularity in recent years, such libraries they have evolved along the technological ladder for the past 30 years. There is lot of interest in digital libraries today. This is reflected in the fact that an advanced Altar Vista search conducted in early July 1996 on “digital library” or “digital libraries” retrieved about 20000 entries.

In spite of the plethora of literature it is not clear what we mean by the term “digital library”. The term is rarely defined, or even characterized. It has been applied to an extraordinary range of applications from digital collaboratories to collection of

electronic journals, software agents that support inquiry based education, collection of e-mail and similar objects, electronic version of a public library, personal library collection and the entire internet among others. It is not easy to see what these have in common except for their digitization. A digital library contains digital representation of the object found in it.

Most understanding of “digital library” probably also assume that it will be accessible via the internet, though not necessarily to everyone. But the idea of digitization is perhaps the only characteristic of a digital library on which there is universal agreement. Digital library is popularly viewed as an electronic version of a library. The term digital library evokes a different impression in various groups. To some it simply means computerization of traditional libraries. But to others who have studied library science, it indicates carrying out the function of libraries in a new way, encompassing new types of information resources, new approaches to acquisition, new methods of storage and preservation, more reliance on electronic systems and networks. But to a computer professional, a digital library is simply a distributed text-based information system, a collection of distributed information service, etc. A digital library is a library of digital documents, artifacts and records. The advantage of having library material in digital form are: (i) the content occupies less space and can be replicated and used electronically, (ii) the content can be made available on networks, (iii) the search for content can be automated.

7.2 DIGITAL LIBRARY: CONCEPT AND DEFINITION

The working group of the US Government’s Information infrastructure Technology and Applications defined the digital library as ‘system’ providing user with a coherent access to a very large organised depository of information and knowledge. R. R. Larson defined digital library as a global virtual library – the libraries of thousands of “networked electronic libraries”. The digital library need not be networked. A digital library is a library which has all the information in electronic form and having electronic devices to have access to the digitized information. Thus digital library is a library which has number of machine-readable publications and facilities for remote access to several databases.

The American Digital Library Federation has defined the digital library as “Digital libraries are organizations 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 collection of digital works so that they are readily and economically available for use by a defined community or set of communities.”

The definition of a digital library can be given as a set of characteristics as follows:

The digital library is :

- a collection of services
- a collection of information objects
- a supporting users with information objects
- organization and presentation of those objects
- available directly or indirectly
- electronic/digital availability

A digital library is much more than just the collection of material in its depositories. It provides a variety of services to all of its users. The basis of the digital library is the information objects that provide the content in the form of digital resource. The goal of the digital library is to satisfy user needs for management, access, storage and manipulation of the variety of information stored in the collection of material that represents the holding of the library. The information objects may be digital objects or they may be in other media but represented in the library via digital means (e. g. metadata). They may be available directly over the network or indirectly. Although the object may not even be electronic, and although the objects themselves may not be available directly over the network, they must be represented electronically in some manner.

Table 1 Describes Essential Properties of a Digital Library Ranging from Quite Traditional to Extremely Broad Views as Set Forth by Harter.

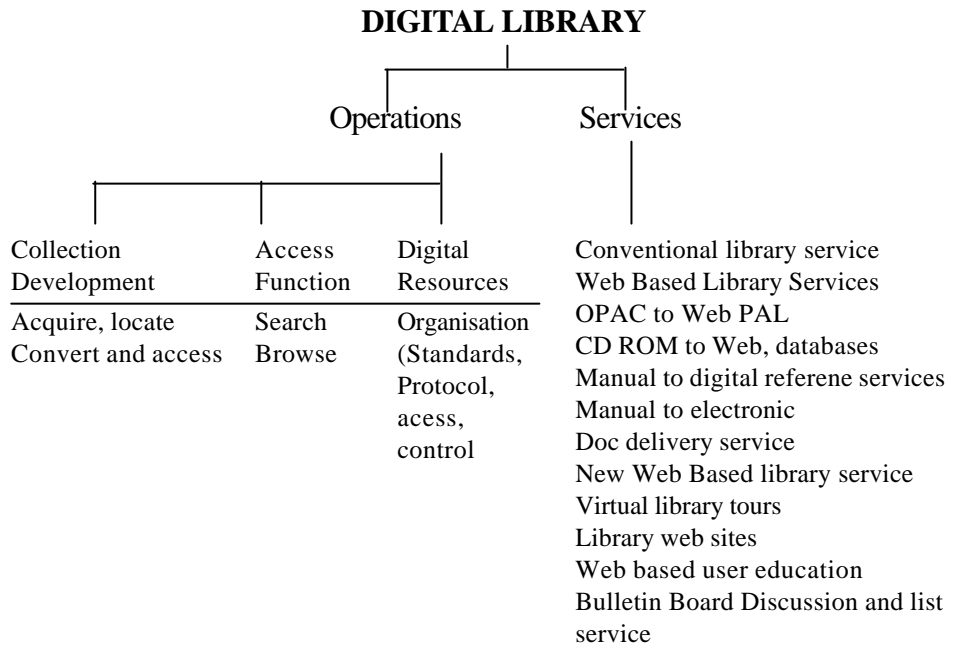
Table 1 : Potential Properties of a Digital Library

Narrow view based on traditional library	Broader view (a model position between the extremes)	Broadest view (loosely based on current network)
Objects are located in a physical place	objects are located in logical place (may be distributed)	Object are not located in physical or logical place
Objects are information resources	Most of the objects are information resources	Object can be anything at all
Objects are selected on the basis of quality	Some of the objects are selected on the basis of quality	No quality control, no entry barriers
Objects are organised		No organization
Objects are subjected to authority control	Some aspects of authority control are present	No authority control
Surrogates of objects are created	Surrogates are created for some objects	No surrogates of objects created
Surrogates are finely searchable	Surrogates and objects are finely searchable	Only object are searchable
Authorship is an important concept	Concept of author is weakened	No concept of author
Objects are fixed (do not change)	Objects change in a standardized way	Objects are fluid (can change at any time)
Objects are permanent	Disappearance of objects is controlled	Objects are transient (can disappear at any time)
Access to objects is limited to specific classes of user	Access to some objects is limited to specific classes of users	Access to everything by everyone

There are many definitions of a digital library. The terms such as “electronic library” and virtual library are often used synonymously. The elements that have been identified as common to these definitions are :

- The digital library is not a single entity.
- The digital library requires technology to link the resources of many
- The linkages between the many digital libraries and information service are transparent to the end user
- Universal access to digital libraries and information services is a goal
- Digital library collections are not limited to document surrogates, they extend to digital artifacts that cannot be represented or distributed in printed formats.

The aim of a digital library may be to expedite the systematic development of digital resources collection; the means to collect, store and organise information and knowledge in digital form.



7.3 CHARACTERISTICS OF DIGITAL LIBRARIES

The structured information in the digital library is called digital object which includes text, audio, video, image, computer programmes, graphics, and multimedia components in digital form. The digital library requires lots of digital technologies. The digital library :

- Provides access to a very large information collection, both primary and secondary
- Support multimedia components
- Provides links to different digital objects
- Supports and provides search and retrieval interface

- Supports the traditional library mission of collection, development, organization, access and preservation.

Important characteristics of a digital library are :

- i) **Digital collection** – In the digital environment a digital library is expected to develop document collection in a digital format.
- ii) **Technology** – It is understood that a digital library will have digital material in its collection. But in the present day context, both digital and non-digital information belonging to a digital library are to be handled using digital technologies.
- iii) **Work and Service** – The professionals supposed to work in a digital library should have necessary training in handling digital information in order to provide the optimum level of effective service.

The most important component of a digital library is the digital collection it holds or has access to. A digital library can have a wide range of resources. It may contain both paper based conventional documents or information contained in computer-processible form. The collection of a digital library may include – a combination of structured/unstructured texts, numerical data, scanned images, graphics, audio and video recordings.

With the assumption that digital libraries are libraries first, some of the important characteristics of a digital library are :

- 1) Digital libraries are the digital face of traditional libraries that include both the digital collection and the traditional, fixed media collection, so they encompass both electronic and paper materials.
- 2) Digital libraries will also include digital material that exists outside the physical and administrative bounds of any one digital library.
- 3) Digital libraries will also include all the processes and services that are the backbone and nervous system of libraries. However, such traditional processes, though forming the basic digital library work, will have to be revised and enhanced to accommodate the differences between new digital media and traditional fixed media.
- 4) Digital libraries will serve particular communities or constituencies as traditional libraries do now, though these communities may be widely dispersed throughout the network.
- 5) Digital libraries will require the skills of both librarians as well as computer professional to be viable.

7.4 DEVELOPMENT OF DIGITAL LIBRARIES

Some of the important points to be considered in developing a digital library are :

- i) Digital collection
- ii) Conversion of existing print into digital format: options for conversion
- iii) Creating portals or gateway to the electronic collection available on the web
- iv) Integrated access interface.

Digital collection

One of the important issues in the creation of a digital library is the building up of a digital collection. Obviously, for any digital library to be viable, it must eventually have a digital collection with the critical mass to make it truly useful. There are essentially three methods of building up a digital collection.

- i) Digitization – converting paper and other media in existing collection to digital form.
- ii) Acquisition of original digital works created by publishers, institutions and other scholars like electronic books, electronic journals and data set.
- iii) Access to external materials not held in house by providing pointers to web sites, other library collections or publishers' services.

While the third method may not exactly constitute part of a local collection, it is still a method of increasing the materials available to users. One of the main issues here is the degree to which libraries will digitise existing materials and acquire original digital works, as opposed to simply pointing to them externally.

A digital library can have a wide range of resources. It can contain both conventional documents and/or information contained in digital or computer-processible form. Rusbridge has divided resources for a digital library into four categories, i.e., legacy, transition new and future. **Legacy** resources are largely non-digital resources including manuscripts, print, slides, maps, audio and video recordings. Attempts are being made to digitise these resources.

Transition resources, are primarily designed for another medium (mostly print) are those which are being or have been digitised, making the transition into the digital world. The conversion into digital form is just to ensure better access and to reduce dependence on physical libraries. These are either digitised images or ways that are converted to list by the process of OCR.

The new digital resources are either deliberately created as digital or are created in parallel to print. Publishers are increasingly moving to XML or SGML format. Future digital resources are – electronic journals, electronic books through databases and data sets in many formats.

The acquisition of documents which are already available in digital formats like CD-ROM databases is a part of the transition. Now-a-days a large number of information products are available on CD-ROM, like MEDLINE, COMPENDEX, METADEX, etc. Libraries can subscribe to these databases (bibliographic or full text) as an important input to the digital collection.

Access to external digital collection

Digital libraries can obtain/acquire access permission to digital collections provided by external sources like other institution, commercial publishers, resources of other libraries, and electronic journals through on-line access. Many of the commercial publishers like Elsevier, Academic Press, ACM, SIAM are making their journals available on-line through web sites. Many other commercial publishers are making available print and electronic versions of their journals. Some of the material is available only in electronic forms.

Conversion of hard copy of the doc into digital format

Now-a-days, the existing selected conventional collection of a library is being converted into digital form. This process of conversion of paper documents into digital format is mainly with the help of scanners. Printed texts, pictures and figures are transformed into computer accessible forms using a digital scanner or a digital camera.

Access to digital information available on web

In the present times, the web provides the hyper media based systems that allow rapid access to a wide variety of networked information resources. One can browse the different web sites which are scattered geographically and have access to the resources. One can download the information. Some of the major portal sites or gateways that provide access to electronic resources are .

[http : // www.edoc.com/](http://www.edoc.com/)

[http : // mel.library.mi.us/](http://mel.library.mi.us/)

[http : // bubl.ac.uk/](http://bubl.ac.uk/)

[http : // sunsite.berkeley.duc/internal/index/](http://sunsite.berkeley.duc/internal/index/)

It can be concluded that digital libraries can develop their collection through the integration of a number of resources and media types. So a digital library may have:

- a) a collection acquired in a digital form
- b) a collection digitised in house
- c) it can purchase access to electronic resources
- d) can gain access to gateways to electronic collections available on the web.

Thus a digital library may not only have a variety of resources but also several mechanism to access these resources. Digital libraries can also provide access to electronic resources through library home page.

Conversion of print to digital

There are basically two methods for converting the print resources into digital resources. These are:

- i) scanning and use of OCR Programmes
- ii) Re-keying of data

Scanning is the easiest option to obtain the exact replica of the original source document. Most scanning software generate default TIFF format which can be converted into PDF format using a number of software. The images can be browsed through a table of contents composed in HTML providing linked to scanned images.

The optical character recognition (OCR) is another method of converting print to digital. The two well-known available OCR techniques are Xerox Textbridge and Caere's omnipage incorporation technology. Retaining the page layout is the major problem in OCR. Now a number of softwares are available for retaining the layout.

The process of retaining the page layout is software dependent. The process of OCR results in computer processible file that is less accurate than re-keying of the data. The acrobat capture 2.0 provides several options for retaining not only the page layout, but also the fonts, and to fit text into the exact space occupied in the original. In image + text solution or OCRed text is generated for each image where each page is an exact replica of the original and left untouched. The OCRed text is generally not corrected for errors since it is used only for searching. The PDF normal gives the clearest image on screen display.

The re-keying (or data entry) or retyping is another mode of converting printed text into digital form. This involves complete typing of the text. This process, although comparatively more accurate, is quite time consuming and costly. But corrections, modifications and changes in formats are easier to carry out.

The four important steps involved in the process of digitisation are scanning indexing, storage and retrieval. The scanning of documents is done with the help of an electronic image scanner. Indexing of a document converted into an image or text file is the second step in the process of document imaging. The most difficult problem is that of storage. The page image has to be stored on a storage media. A large capacity storage device has to be used, like optical disc- CD-ROM, magnetic tapes, etc.. After the storage of scanned image is the retrieval of the needed data from the digital collection files. The digital image is also called “bit mapped page image”.

The technical requirements of the digital imaging process (i.e. converting paper documents into digital images) to be made accessible over network include :

- Hardware (scanners, computer, data storage and data output peripherals)
- Software (image capturing, data compression)
- Network (data transmission)
- Display technologies

Tools of Digitisation

The important machines and tools needed for digitisation include :

- i) Computers
- ii) Scanners and scanning software
- iii) Storage System
- iv) Network
- v) Display system
- vi) Printer

Scanners – Some of the important scanners being used for capturing digital images are:

- i) HP Scanjet 6300 C
- ii) Ricoh IS420
- iii) Bell and Howell 1000 FB
- iv) Kodak 500S

- v) Juno CP-4000
- vi) Digital camera – Zentschel Omini Scan 3000, Minolta PS3000
- vii) Slide scanner – Kodak PCD scanner 4045, Nikon LS3510
- viii) Microfilm scanner – Mekel M500 XL, Sunrise. SR1-50, Lenzpro 2000 Multimedia digitiser.

Some of the image capturing/scanning software are : Altris software, OPTIX, Documentum, Power office, File net, Java system, Docs open.

7.5 DIGITAL LIBRARIES AND THEIR USES

The important functions and uses of the digital library in context of users are that it :

- i) provides access to a very large information collection in a digital form
- ii) supports multi-media content
- iii) is network accessible
- iv) provides user friendly interface
- v) offers links to local/external objects
- vi) support advance search and retrieval
- vii) supports the traditional library mission of collection, development, organization, access and presentation
- viii) supports publishing, annotation and integration of new information
- ix) brings together people with formal, informal and professional learning missions
- x) provides faster access to information resources
- xi) provides an easy mechanism for resource sharing with other libraries. Sharing of digital files is much easier.

7.6 MAJOR ISSUES/CHALLENGES

Creating effective digital libraries poses serious challenges. Some of the more serious issues facing the development of digital libraries are :

- i) **Technical architecture** – Libraries need to enhance and upgrade current technical architecture such as:
 - High speed local network and fast connection to internet
 - Relational database that supports a variety of digital formats
 - Full text search engines to index and provide access to resources
 - A variety of servers such as web services and FTP servers
 - Electronic document management system
- ii) **Building digital collections** - One of the most important issues in creating a digital library is building of the digital collection. One of the major issue is the degree to which libraries will digitise existing material and acquire original digital

works. This is the old access versus ownership issue. How is the specific material to be digitised / to be acquired to be identified by a given library. Who collects and/or digitises which material could be based on factors such as – collection strength, unique collections, the priorities of user groups, manageable portions of collection, technological resources and skills of the staff.

iii) **Digitisation** : Another aspect is what portion of collection to digitise. Digitisation is conversion of any fixed or analogue media – such as books, journal articles, photos, paintings, microfilm into electronic form either through scanning or rekeying. There are several approaches available, at least theoretically.

- Retrospective conversion of collections
- Digitisation of a particular special collection or a portion of one
- Highlight a diverse collection
- High use materials
- An ad hoc approach –(one digitises and stores material as they are requested)

iv) **Metadata** – Metadata is the data that describe the content and attributes of any particular item in a digital library. Metadata is important for digital libraries because it is the key to resources, discovery and use of any document like a library catalogue. The “Dublin Core” is one of the prominent schemes. There is the problem of naming, identifiers and persistence. Naming is required to uniquely identify digital objects. Any system of naming should be permanent, lasting indefinitely. The names cannot be bound with specific locations. A global scheme of unique identifier is required. Three schemes proposed to get over the problems of persistent naming are PURLs, URNs, and digital object identifiers:

PURLs – are persistent URLs, a scheme developed by OCLC to separate document name from its a location.

URN – Uniform Resource Name have been developed by Internet Engineering Task Force (IETF).

Digital Object Identifier (DOI) – Developed by Association of American Publishers and Corporation for National Research Initiatives to provide a method by which digital object can be reliably identified and accessed .

v) **Copyright/rights management** - Copyright is one of the most important barriers to digital library development. The current paper- based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects are less fixed, easily copied and remotely accessible by multiple users simultaneously. The problems of libraries are that they are for the most part simply caretakers of information, they do not own the copyright of the material they hold. So libraries will never be able to freely digitise and provide access to the copyrighted material in their collection. They have to develop a mechanism for managing copyright.

vi) **Preservation** – Another important issue is preservation. In the preservation of digital material, the real issue is technical obsolescence. There are three issues of preservation:

- Preservation of the storage medium - tapes, hard drives, floppy discs — have a short life span when considered in terms of obsolescence.

- Preservation of access to content – this form of preservation involves preserving access to the content of the document regardless of the format. While files can be moved from one storage medium to another, what happens when the formats (e.g., Acrobat PDF) containing the information becomes obsolete?

Self Check Exercise

- 1) Define digital library.
- 2) What are the characteristics of a digital library?
- 3) What are the major issues/challenges in the development of the digital library?

Note : i) Write your answers in the space given below.

ii) Check your answers with the answer given at the end of this Unit.

.....

.....

.....

7.7 SUMMARY

There is much confusion with regard to the definition of the “digital library”. There is such confusion surrounding the terms partly because library professionals have been using different phrases over the years to denote this concept – electronic library, virtual library, library without walls, etc. Another factor adding to the confusion is that digital libraries are at the focal point of many different areas of research. The third confusion arises from the fact that there are many things on the Internet that people are calling “digital libraries” but which differ from the library professionals point of view. Many people consider World Wide Web as a digital library.

Creating effective digital libraries poses a serious challenge. The integration of digital media into traditional collection will not be straight forward because of the unique nature of digital information — it is less fixed, easily copied and remotely accessible by multiple users simultaneously.

7.8 ANSWERS TO SELF CHECK EXERCISES

- 1) A digital library is defined as a library which has all the information in electronic form and having electronic devices to have access to digitised information. The digital library is a collection of services, a collection of information objects, which supports users with information objects. Access to digital resources is made directly or indirectly and ensures electronic digital availability. The American Digital Library Federation has defined digital library as an “organizations that provide the resources, including the specialised 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.”

- 2) The important characteristics of a digital library are :
 - i) Digital collection – the digital collection is the main resource of a digital library. It may include both digital collection as well as traditional collection.
 - ii) Digital libraries also include digital material that exist outside the physical and administrative bounds of any one digital library.
 - iii) It includes all the processes and services that are the backbone of libraries.
 - iv) It serves particular communities or constituencies as traditional libraries do, though these communities may be widely dispersed throughout the network.
 - v) It requires both the skills of the librarian as well that of the computer specialist to be viable.
 - vi) Supports multi-media components.
- 3) The major issues in the development of a digital library are the following :
 - i) Technical architecture
 - ii) Building digital collection
 - iii) Digitization
 - iv) Metadata
 - v) Copyrights/rights management
 - vi) Preservation.

7.9 KEYWORDS

- Digital Library** : Library with both digital collection and traditional, fixed media collection. It also includes digital materials that exist outside the physical and administrative bounds of any one digital library.
- Digital Collection** : It stands for texts, images, databases, files, etc. in a digital format. It comprises documents which are in digital formats like CD-ROM Database , floppies, magnetic tapes, discs, etc.
- Digitisation** : It is the process of converting print information and other media information of the existing collection to digital form. It is the conversion of any fixed or analog media, such as books, journal articles , photos, paintings, microforms, etc., into electronic form through scanning or in fact even re-keying.
- Metadata** : It is the data that describes the content and attributes of any particular item. It is the key to resource discovery and the use of any document. The cataloguing record (i.e. bibliographical record) is an example of metadata.

- Scanners** : It is a device used to capture a digital image from an analogue media such as the printed page or a microfiche/microfilm at a predefined resolution.
- Scanning** : The process of acquisition of an electronic image through its original that may be a text, photograph, manuscript, etc., into the computer using an electronic image scanner. An image is read or scanned at a predefined resolution and dynamic range.
- Re-keying** : The action of entering data into a computer using a keyboard.
- Copyright** : The legal right exclusively given for a definite period of time to the originator (authors or creator) of intellectual work such as a publication, or a literary work. The legal issues of electronic information include copyright, ownership, pricing, rules and regulations governing multiple usage.

7.10 REFERENCES AND FURTHER READING

- Arora, Jagdish (2001). Building Digital Libraries : An Overview. *DESIDOC Bull of Information Technology*. 21; 3-24.
- Harter, Stephen P (1997). Scholarly Communication and the Digital Library : Problems and Issues. *Journal of Digital Information*. 1; 4-14.
- Army, W Y (1995). Key Concepts in the Architecture of the Digital Library. *D-library Magazine*. July. URL : <http://www.dlib.Organization/July 95/07 arms.html>.
- Chapman, S and Kenny A R (1996). Digital Conversion of Research Library Materials. A Case for Full Informational Capture. *D-Library magazine*. Oct 1996. URL : <http://www.dlib.organization/dlib/october96/10 chapman/html>.
- Lest, M (1996). Going Digital. *Scientific American*. March 1996, 58-60.
- Schatz, Bruce R (1997) Information Retrieval in Digital Libraries : Bringing Search to the Net. *Science* . 275; 327-33.
- Smith, T.R. (1997). Meta Information in Digital Libraries., *Int J. Digital Libraries*. 1; 105-7.
- Fox, E A and Lumin, L (1993). Introduction and Overview to Perspectives on Digital Libraries. *JASIS* 44(8); 441-443.
- Kenney AR and Chapman S. (1996) Digital Imaging for Libraries and Archives. NY: Cornell University, 198 p.
- Noerr, P (2000). The Digital Library tool Kit; Ed. 2 Palo Alto : Sun System, 186 p.
- Malinconico, SM and Warth JC (1996). Electronic Libraries : How Soon. *Program*. 30(2); 133-148.
- Marcuum Deanna, B (1997). Digital Libraries : for whom ? for what ? *J of Academic Librarianship*. 23(2); 81-84.