DEVELOPING DIGITAL LIBRARY

BS(LIS)

Code No. 9216 **Units: 1-9**

Department of Library and Information Sciences
Faculty of Social Sciences and Humanities
ALLAMA IQBAL OPEN UNIVERSITY
ISLAMABAD

DEVELOPING DIGITAL LIBRARY

BS LIBRARY AND INFORMATION SCIENCES

Course Code: 9216 **Units: 1-9**

Department of Library and Information Sciences
Allama Iqbal Open University, Islamabad

COURSE TEAM

Chairman: Dr. Pervaiz Ahmed

Compiled by: Muhammad Jawwad

Reviewed by: Dr. Amjid Khan

Program Coordinator: Muhammad Jawwad
Lecturer

Course Coordinator: Muhammad Jawwad
Lecturer

Editor: Humera Ejaz
Fazal Karim

Layout: Naeem Akhtar

FOREWORD

Department of Library and Information Sciences established under the flagship of the Faculty of Social Sciences and Humanities with the aim to produce trained professional graduates. Since its establishment the department is offering various programs from certificate level to PHD level. The department is supporting the mission of AIOU keeping in view the philosophies of distance education. The primary focus of its programs is to provide quality education through targeting the educational needs of masses at their doorsteps all over the country.

BS 4-year program is a competency-based learning program. The primary aim of this program is to produce knowledgeable ICT-related skilled professionals. This program scheme of study is especially designed on the foundational courses which provide in-depth knowledge and understanding of the area of specialization. It also focuses on the general subjects and theories, principle, and methodologies of related LIS and relevant domains.

This new program has a well-defined level of knowledge which includes general education courses, foundational skills courses. The students are expected to advance beyond their secondary level, and mature and deepen their competencies, including in writing, communication, mathematics, languages, analytical and intellectual discipline. Moreover, the salient feature of this program is its practical learning requirement component which provides students a platform of practical knowledge of the environment which they will face in future of their professional life.

This program intends to enhance student's abilities in planning and controlling library functions. The program will also produce highly skilled professional human resource to serve the libraries, resource centers, documentation centers, archives, museums, information centers, and LIS schools. Further, it will also help the students to improve their talent of management, research, Technology, advocacy, problem-solving and decision making relevant to information work in a rapidly changing environment along with integrity and social responsibility. I welcome you all and wish you good luck for your academic exploration at AIOU.

Prof. Dr. Zia Ul Qayyum
Vice Chancellor

PREFACE

A library is a place where information is disseminated professionally to users in a user-friendly environment. The application of computers to various tasks by which information is collected, processed, organized and showcased was a significant development in the later part of the last century/Computerized processes gave rise to dynamic as well as interactive products such as the inventory/database of the collection that can be searched and accessed electronically, e.g., the online public access catalogue (OPAC), in contrast to the comparatively static card catalogue. The single most important development that has brought about sweeping changes in the library and information discipline in the developed world is that of digital libraries. Though the professionals and libraries in developing countries are also experiencing the virtues of the Internet and electronic information highways, many of these libraries have not gone much farther than the computerization of in-house operations, making use of databases available in electronic media such as CD-ROMs, and Web access to subscribed journals and various free resources. Some of these libraries created websites basically to project the services and strengths of the library and to serve more as advertising or public relations media rather than digital information gateways. Digital library development should be taken up as an additional task to populate the websites with valuable in-house content like research reports, publications of inhouse researchers, and so on. In a country such as Pakistan, which is so rich in indigenous research and development in disciplines varying from science and technology to social science, humanities and spirituality, there is a tremendous need for hosting full-fledged digital libraries by appropriately tagging this content with affordable information technology. The recent years witnessed the boom of the Internet the world over, leading to the acceptance of the Web as an alternative delivery mode of information products. Currently we rarely see a renowned publisher without a website, and most of the international journals offer some provisions to access the abstracts/full text of papers through the Web along with the print subscription. The information center is undergoing a transition from the paper-dominated environment to the shared access-oriented electronic environment. Digital libraries as a fresh, lively and dynamic area with a lot of enthusiasm and activity by researchers from different disciplines, institutions and countries. The Information Infrastructure Technology and Applications (IITA) Working Group considers ‘digital libraries as systems providing users with coherent access to a very large, organized repository of information and knowledge.’

This course is planned for use in both graduate programs of library and information science and the undergraduate and technical programs leading to degrees and certificates a career in library services. It has also been widely used for training purposes within various library and information science departments to train

students and library professionals. The main aim of this course is to give a broad-based view of various functions of Digital Library Development. Here in this course, we introduce many functions of digital library, models of digital library, planning techniques of digital library and the duties and responsibilities of both professional librarians and the support staff, which can overlap in different types of libraries, but the main function is the same.

The course provides both an historical and contemporary perspective of digital library movement and discusses issues such as sustainability and preservation. It also provides a practical approach to digital libraries and present a comprehensive picture of digitization and explains the process of digitization in a step-by-step approach along with different file formats and alternatives to digitization. The contents discussed in the course can form the models for action practical approaches to electronic records management and preservation. It also presents practical tools that seamlessly integrate into the system design process and result in the identification of technical specifications and opportunities for improving performance through improved access to records.

The contents of course will also explain digital library's role in the context of e-governance and will explain the concept of digital divide and provides a brief overview of digital divide and the effects of some critical factors on it. Unequal investment of Information and Communication Technologies (ICTs), the potential of the Internet, the important role of education, literacy, education, e-governance, librarians, libraries and also digital libraries etc. are some factors discussed in this unit. It also discusses that paying attention to all so-called critical factors can bridge and decrease this global digital divide and its present situation in Asia-pacific countries. Finally, the course will explain the development of digital library using open-source software DSpace. DSpace is an open source software platform that enables organizations to: capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options distribute an organization's digital assets over the web through a search and retrieval system preserve digital assets over the long term, discuss a functional overview of the system, which is a good introduction to the capabilities of the system, everyone student should read this unit because it introduces some terminology and make familiar with the DSpace in a practical way: Functional overview, Installing DSpace, Upgrading DSpace, Using DSpace, Development in DSpace etc.

Prof. Dr. Syed Hassan Raza
Dean
Faculty of Social Sciences and Humanities

ACKNOWLEDGEMENTS

All praise to Almighty Allah who has bestowed me the potential and courage to undertake this work. Prayers and peace be upon our Prophet Hazrat Muhammad, his family and all of his faithful companions.

I am thankful to the worthy Vice-Chancellor and the worthy Dean, FSSH for giving me the opportunity to prepare this study guide. Without their support, this task may not be possible. Further, they have consistently been a source of knowledge, inspiration, motivation, and much more.

I am highly indebted to my parents, spouse, siblings, and children, who allowed me to utilize family time in completion of this work timely. Their continuous prayers kept me consistent throughout this journey. I would also appreciate the cooperation of my departmental colleagues extended to me whenever required. Special thanks to Academic Planning and Course Production (APCP) and Editing Cell of AIOU for their valued input that paved my path to improve and finish this study guide in accordance with AIOU standards and guidelines. They have been very kind and supportive as well.

I would also like to thank Print Production Unit (PPU) of AIOU for their support with regard to comprehensive formatting of the manuscript and designing an impressive cover and title page. Special thanks also owes to AIOU's library for giving me the relevant resources to complete this task in a befitting manner. I am also thankful to ICT officials for uploading this book on AIOU website. There are many other persons, whose names I could not mention here, but they have been a source of motivation in the whole extent of this pursuit.

Muhammad Jawwad
Course Coordinator

OBJECTIVES OF THE COURSE

After completion of this course the students will be able to:

1. Understand the concept of digital library and its various aspects?
2. Familiar with various methods, tools, techniques and technologies of digitization and digital conversion.
3. Explain the role of open-source software and different approaches of building digital libraries in the context of web 2.0 technology.
4. Verbalize information preservation challenges, and information preservation and information services in the digital age.
5. Mange reference and information services, and digital repositories in digital environment.
6. Explain the role of digital library and e-governance in context to digital divide in developing countries.
7. Develop the digital library using DSpace open source software.

Recommended Readings:

1. Ashraf, T., Sharma, J., & Gulati, P. A. (2010). *Developing sustainable digital libraries: Socio-technical perspectives*. Hershey: Information Science Reference.

2. DSpace 5.x documentation (2015). Retrieved November 15, 2015, from https//wiki. duraspace.org/display/DSDOC5x

COURSE ORGANIZATION

The course has been designed as easy as possible for distance mode of learning and it will help students in completing his/her required course work. The course is of three credit hours and comprises on nine units, each unit starts with an introduction which provides an overall overview of that particular unit. At the end of every unit the objectives of unit shows student the basic learning purposes. The rationale behind these objectives is that after reading unit a student should be able to explain, discuss, compare, and analyze the concepts studied in that particular unit. This study guide specifically structured for students to acquire the skill of self-learning through studying prescribed reading material. Studying all this material is compulsory for successful completion of the course. Recommended readings are listed at the end of each unit. Few self-assessment questions and activities have also been put forth for the students. These questions are meant to facilitate students in understanding and self-assessment that how much they have learned.

For this course, a 3-days workshop at the end of semester, and four tutorial classes/meeting during semester will be arranged by the department for learning this course. The participation/attendance in workshop is compulsory (at least 70%). The tutorial classes/meetings are not formal lectures as given in any formal university. These are meant for group and individual discussion with tutor to facilitate students learning. So, before going to attend a tutorial, prepare yourself to discuss course contents with your tutor (attendance in tutorial classes/meetings is non-compulsory).

After completing the study of first 5 units the 'Assignment No. 1' is due. The second assignment that is 'Assignment No. 2' is due after the completion of next 4 units. These two assignments are to be assessed by the relevant tutor/resource person. Students should be very careful while preparing the assignments because these may also be checked with Turnitin for plagiarism.

Course Study Plan and Chart

As you know the course is offered through distance education so it is organized in a manner to evolve a self-learning process in absence of formal classroom teaching. Although the students can choose their own way of studying the required reading material, but advised to follow the following steps:

Step-1: Thoroughly read description of the course for clear identification of reading material.

Step-2: Read carefully the way the reading material is to be used.

Step-3: Complete the first quick reading of your required study materials.

Step-4: Carefully make the second reading and note down some of the points in note book, which are not clear and needs fully understanding.

Step-5: Carry out the self-assessment questions with the help of study material and tutor guidance.

Step-6: Revise notes. It is quite possible that many of those points which are not clear and understandable previously become clearer during the process of carrying out self-assessment questions.

Step-7: Make a third and final reading of study material. At this stage, it is advised to keep in view the homework (assignments). These are compulsory for the successful completion of course.

Assessment/Evaluation Criteria of Students' Coursework

As per AIOU's policy.

Note: Assignments submission and getting pass marks is compulsory, the student who will not submit assignments or marked as fail considered FAIL in the course. He/she will need to get fresh admission in the course.

CONTENTS

Unit. 1

DIGITAL LIBRARIES: AND INTRODUCTION

Complied by: Muhammad Jawwad

Reviewed by: Dr. Pervaiz Ahmed

CONTENTS

INTRODUCTION

This unit is developed to establish a background of the topic, it will inhabit the concept, development, issues, and achievements of digital libraries in students. It will also discuss the characteristics, role and features of digital library and challenges in learning the digital library.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is digital library?
2. What are key components and features of digitization and the characteristics of digital library?
3. What is digital imaging and advantages of digital contents?
4. What challenges libraries face in digital preservation?
5. What are different digital library services?
6. How digital libraries play a role in the provision of scholarly communication in an electronic format?

1.1 Introduction

Numerous terms are used by authors to define the concept of digital libraries. According to E.A. Fox the digital library may be defined as the "New way of carrying out the functions of libraries encompassing new types of information resources, new approaches to classification and cataloguing, intensive use of electronic systems and networks and dramatic shifts in intellectual, organizational and electronic practices".

Larsen defines, "A digital library as a global virtual library the library of thousands of networked electronic libraries".

The Associations of Research Libraries (ARL) (1955) has identified five elements common to all definitions of digital library as follows:

1. The digital library is not a single entity,
2. The digital library by requires technology to link resources of many,
3. These links are transparent to end-users,
4. Universal access to the digital libraries and information services in the goals, and
5. Digital library collections are not limited to document surrogates, but they also include digital artifacts that cannot be represented or distributed in printed formats.

In broad sense digital library is a computerized system that allows users to obtain a coherent means of access to an organized, electronically stored repository of information and data. The digital library is on electronic library consisting of information in the digital, analog and digitized form.

Librarians tend to speak for a broader definition of the term "Library". They see a library as an organisation that secures the selection, conservation, organisation, preservation and the access to information that is vital for the members of the specific organisation. Librarians and Libraries carry a long history and tradition that has been somewhat cemented during the centuries. With the coming of the Internet and digital media, for librarians this is only yet another delivery channel for yet another media.
Researchers most often favour a narrower definition of the library concept. For them a library could be any room containing a small or large amount of books or data discs or tape cassettes. Researchers seldom care for the social and institutional context of the term "Library". Their emphasis is tilted towards databases and how to collect, retrieve, organise and access the information.

But through the 1990s definitions of Digital Libraries have broadened in scope even if there is no definitive one. Trying to summarise definitions given by Research initiatives, Science Foundations and Digital Library researches, three elements seems to be necessary:

1. There must be some sort of organised collection.
2. It can be partly bibliographical but full-text files of the data, if it is an article or manuscript etc., is now frequently added and required in various formats.
3. The collection is organised for a group or community of users.

In 1998 Donald J. Waters presented the first, short and workable librarians definition of Digital Libraries, which also was adopted by the Digital Library Federation (DLF) who´s members are major American universities as well as Library of Congress and British Library. It reads:

> "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 collections of digital works so that they are readily and economically available for use by a defined community or set of communities."

Here in this definition the focus is on the Digital Library as an organisation providing information services in digital form and also taking responsibility for preservation and integrity of the collection. The definition is much broader. But the fact that a library is calling itself for a Digital Library does not usually mean that all its services are digital. It usually means that some parts of the information services are digital. Very few libraries are Digital Libraries in the sense that their services are digital only.

1.2 Characteristics of Digital Libraries

Digital/digitized libraries require digital technologies and is the combination of traditional and media collection. So, they encompass both paper and electronic material. Digital library consisted of mostly electronic documents, which are of reference type.

- Digital Library is a digital object, which may be text, audio, video, image, and numeric multimedia components.
- Digital library can be accessed from user's workplace.
- Digital libraries support formal and informal learning procedures.
- Digital library provides remote to rare and expensive
- material.

1.3 Digitization: Key Components and Features

A fully developed digital library environment involves the following key elements:

1. Initial conversion of content from physical to digital form.
2. The extraction or creation of metadata or indexing information describing the content to facilitate searching and discovery, as well as administrative and structural metadata to assist in object viewing, management, and preservation.
3. Storage of digital content and metadata in an appropriate multimedia repository. The repository will include rights management capabilities to enforce intellectual property rights, if required. E-commerce functionality may also be present if needed to handle accounting and billing.
4. Client services for the browser, including repository querying and workflow.
5. Content delivery via file transfer or streaming media.
6. Patron access through a browser or dedicated client.
7. A private or public network.

1.4 Some Advantages of digital contents

There are number of advantages of digital contents some of them are as follows:

1.4.1 Easy Manipulation: Digital objects, by their very nature, can be easily stored, recalled, and manipulated. Current photo editing software allows digital images to be resized, duplicated, and edited using techniques simply aren't possible using the established print or analog methods for image processing.

1.4.2 Easy Access: The most important advantage of digital objects, however, is accessibility. Digital documents can be easily posted on web sites and online resources, allowing anyone with Internet access the ability to retrieve them. The storage of images and documents in digital format, and making them accessible through digital libraries, allows numerous simultaneous users the ability to view the same documents at once, while eliminating analog duplication costs.

1.4.3 Interactive Learning: Digital image and text can include videos and other multimedia content, and links to related websites, to make the learning experience more productive and more engaging.

1.4.4 Study Smarter: Digital contents and images are the smarter study concept of libraries. User can highlight and add notes to contents where feel necessary.

1.4.5 Simple Navigation: Navigation and exploration in the digital contents are simple and easier than analog. Search keywords and phrases throughout all of the digital contents or quickly jump to a location by typing a page number is a smarter feature of digital collection.

1.5 Digital Imaging: Technological Issues

As the digital libraries become more heavily involved in projects that require the digitization of documents and images, it becomes increasingly necessary to ensure that everyone involved in such projects have a working knowledge of digital imaging technology. This is not an easy task and requires as much hands-on experience as it does knowledge of the theories and concepts behind digital imaging.

A digital image is little more than a file which contains instructions to appropriately illuminate and color a group of pixels, in order to create what the human eye sees as an image. Each pixel can be made any of up to 16 million colors, by varying the levels of red, green and blue light that illuminate each pixel. A digital image is produced using a similar method. All computer displays divide a screen into a number of pixels, or digital image elements.

1.5.1 Digital Imaging General Principles

When digitizing documents, it is important to consider possibilities beyond the scope of the immediate project. The goal is to eliminate as much as possible the prospect of having to repeat steps, and re-digitize a document whenever a new need with different requirements arises. Without such a plan in place, one may find himself re-scanning the same material as a resultof poor image quality, inability to meet changing project requirements, or obsolescence of the technology being used. While such possibilities can't be eliminated, we can reduce the likelihood of such an occurrence by implementing a plan for scanning at high resolutions, in open formats.

1.5.2 Scanning Resolution: The image resolution determines how many pixels per inch are used to define an image. The higher the resolution, the more definition and detail can be stored in an image, permitting greater options in enlargement and manipulation.

1.5.3 File Types: A number of standards exist for storing digital images. Each have advantages and disadvantages in terms of image quality, storage space, and compression capability.

- TIF: Best quality for master and archive files.
- TIF without LZW compression: Gives maximum compatibility with PC, Mac, UNIX etc, while being a lossless compression.

- JPEG: Best quality for use copies, good resolution, small file size. Works best with photographs and other continuous-tone images.
- GIF: Small file but limited in color hence is not good for photographic images. Works best with images composed of lines and solid blocks of color, such as text, cartoons, or buttons.
- PDF: Displays the exact same content and layout no matter which operating system, device or software application it is viewed on. It is multi-dimensional format allows us to integrate various types of content – text, images and vector graphics, videos, animations, audio files, 3D models, interactive fields, hyperlinks, and buttons. All these elements can be combined within the same PDF file and organized as a report, a presentation or a portfolio. It is easy to create, read and use by everyone and secure.

As digital libraries scale in size and functionality, it is critical for the underlying technology platform to deliver the performance and reliability required. Patrons expect high service levels, which means that downtime and poor response time are not tolerable. Moreover, because cost is a foremost concern, scalability and efficiency with a low total cost of ownership are also key requirements. This type of digital library implementation requires a scalable enterprise-level technology solution with built-in reliability, availability, and serviceability (RAS) features.

Storage capacity also must be scalable to adapt to rapid growth in demand, and must be adapted to the mix of media types that may be stored in a digital library, such as:

- Text, which is relatively compact.
- Graphics, which can be data intensive.
- Audio, which is highly dynamic.
- Video, which is highly dynamic and data intensive.
-

Storage capacity should be expandable in economical increments and should not require redesign or re-engineering of the system design as requirements grow. An open systems architecture provides both a robust platform and the best selection of digital media management solutions and development tools. The inherent reliability and scalability of open platforms have made them the most popular choice of IT professionals for Internet computing.

This computing model features an architecture that is oriented totally around Internet protocols and stresses the role of Web sites for a vast and diverse array of services that follow a utility model.

1.5 Digital Preservation

For Preservation of digital materials on large scale and as per the storage capabilities and in standard formats that are accessible and usable, it is necessary to articulate some basic requirements. Digital preservation requirements needs to be examined both from the perspective of users of digital materials and from the view of libraries, archives, and other custodians who assume responsibility for their maintenance, preservation, and distribution as separate but integrated stakeholders.

To achieve their preservation goal and satisfy the requirements of their users, librarians need to preserve materials in formats that enable the types of analyses that users wish to perform despite resource constraints. By making explicit preservation requirements from both the users' and custodians' perspectives, libraries and archives will be better able to integrate digital preservation into overall planning and resource allocation.

The use of digital materials has immense potential but has a varied and unpredictable nature. Accordingly any generalization, even if restricted to one community of users such as humanities scholars would be hazardous and runs the risk of overlooking and understating potential user needs. Exact needs for presentation and analytical tools may have a variation among disciplines, yet some intrinsically basic requirements are likely to transcend fields and disciplines. "The ability to establish the authenticity and integrity of a source is critical to users, whether it is generated by an individual, created in the conduct of institutional business, or produced through a formal publication process".

System that will facilitate establishment of authenticity of an object require archives and libraries to store much more than the content of digital documents. According to Graham (1995), "attributes such as formal document structures, metadata that document the maintenance and use history of the document, time and date stamps, and a series of references among documents are essential for determining authenticity and for understanding the provenance of sources and placing them in a larger context." Largely depending upon requirements of users, availability of sufficient funds and technical be expressed differently by libraries, and other types of repositories that are struggling to meet.

Storage systems should be capable of handling digital information in a wide variety of formats, including text, data, graphics, video, and sound. Digital storage both as an only and alternative means for storing print formats should have a long life expectancy, a high degree of disaster resistance, sufficient durability to withstand regular use, and very large storage capacities. Conversion from analog to digital formats and migration to new generations of technology should permit very large scale transfers of heterogeneous materials. Developing and sustaining digital library is an expensive and resource-intensive process. Hence it is important to

consider some basic principles underlying the design, implementation, and maintenance of any digital library. These principles apply across the spectrum covering not only projects in which analog objects are converted to digital form, but to mixed digital libraries in which the objects may be of both types.

Changing technologies can quickly outperform the ability of designers to maintain a particular digital library. We must be able to anticipate and plan for changes needed to provide lasting access to its information. Since most digital library projects are long-term efforts, they require the commitment of long-term resources. Beginning such a project involves an implicit, if not explicit, commitment to the continuation of the work and a promise that the digital materials will continue to be available. Lacking organizational commitment, it may not make sense to even begin a project.

1.7 Challenges of Digital Preservation

Librarians have been instrumental in developing wide ranging tools and methodologies to reduce the decay of traditional materials and to restore books and documents that have deteriorated to such an extent that their longevity and usability are threatened. Much remains to be done to preserve cultural, intellectual, and scholarly resources in traditional formats that form the foundation for humanities research and teaching.

According to a research conducted in 2004 an estimated 80 million embrittled books reside in American libraries, 10 million of which are unique; and countless journals, newspapers, photographs, and documents require preservation treatment to survive well into the next century.

However, preservation of digital objects is a problem and challenge of an altogether different kind which is quite different from preserving traditional format materials. Preservation and guaranteed maintenance of digital materials is much discussed. Whereas the preservation of printed materials harbours several specific and well-known problems, these issues differ from those encountered with digital materials, where there exist many technical issues, ultimately solvable by technical means but with few standard solutions available, as yet.

Although librarians have been fighting acid-based papers, thermo-fax, nitrate film, and other fragile media against deterioration since long, the preservation challenge posed by magnetic and optical media is altogether of different kind as they can deteriorate rapidly, making the time frame for decisions and actions to prevent loss is a matter of years, not decades. Since advances and innovation in the computer hardware, storage, and software industries continues at a fast pace, generating greater storage and processing capabilities at reduced cost, more than media deterioration is the problem of obsolescence in retrieval and playback technologies.

Devices, processes, and software for recording and storing information are being replaced with new products and systems both unexpectedly and on a regular two- to three-year cycle, driven primarily by innovation and market forces rendering records prepared in digital form in the first instance and those converted retrospectively from paper or microfilm to digital form equally vulnerable to such a fast technological obsolescence.

The absence of well established standards, procedures, and well tested methods for preserving digital information has further aggravated the problem of long term preservation. Digital library research has largely focussed on architectures and systems for information organization and retrieval, presentation and visualization, and administration of intellectual property rights. The important role of digital libraries and archives in ensuring the future accessibility of information with enduring value has taken a back seat to enhancing access to current and actively used materials.

1.8 Digital Library Services

Deployment of portal technology has occurred side-by-side with digital collection development and access. As libraries create, license, or negotiate access to more and more digital content, the need for an easy-to-use interface becomes increasingly important. Library portals typically include an online catalogue of materials as well a gateway to collections of digital resources accessible to the user.

Broadcast search tools allow library users to search all of these sources simultaneously with a single query. Portals may include electronic reference services (“ask a librarian”), personalization features (“my bookshelf,” custom intelligent search), and other research tools. Enriched content, such as author biographies and book reviews, tables of contents, and jacket images can be provided to supplement the online catalogue. Some libraries have built interactive features into their portals, allowing development of virtual communities.

1.8.1 Generic Digital Library

The University of California at Berkeley has been in the forefront of digital library innovation for many years. Projects begun at Berkeley included the creation of specifications for encoding electronic finding aids that are used to access special collections and archives. These specifications later evolved into the XM-based Encoded Archival Description (EAD). Current work involves development of a modular object management environment, called Gen DL (for “Generic Digital Library”) which can be accessed from the website: https://cdlib.org/. Gen DL includes three components that can be thought of as three separate systems

1. The Web-based content management system, where creation and maintenance of the digital content is controlled. Descriptive, administrative, and structural metadata is created and linked to digitized or born digital content. The resulting digital library objects are then encoded to the METS standard.
2. The preservation repository (a joint project with the California Digital Library), where the digital content is managed to ensure its integrity and longevity.
3. The access system, which is used to discover, display, and navigate objects that may have complex internal organizations (i.e., structural metadata).

1.8.2 Cornell Digital Library

Having undertaken one of the world's most aggressive digitization programs, Cornell library has been a developer of best practices and a leader in the field of digital libraries. As a Sun Center of Excellence for Digital Libraries, Cornell has served as a co-developer with Endeavor Information Systems for the product "ENCompass," a specially tailored multimedia management system intended for library implementations. Although Cornell has gained international recognition for its computer science advances in digital library technology, this university also recognizes the value of off-the-shelf products can be accessed (http://www.library.cornell.edu/)

Cornell has attained recognition for the substantial digital collections it has developed. One of the best known is the "Making of America" collection, a multi-institutional digitization initiative making available primary sources of 19th century.

American culture and history, including popular magazines like Harper's, Atlantic Monthly, and Scientific American. To date, more than 900,000 pages, which are full-text searchable and freely accessible over the Internet, are in the collection.

More than 40 other digital collections have been established, ranging from audio recordings of rare birdcalls to digital facsimiles of essential works in the literature of witchcraft and demonology, drawn from Cornell University's Rare Book Collection.

There are several strategic digital library priorities. One is to establish a central repository for all information resources deemed worthy of long-term maintenance. The repository's aim is to support the total lifecycle of those materials.

1.8.3 Harvard University Library Network

Harvard University has a very large and decentralized library system including more than 100 libraries and 35 major research collections. There is no hierarchy tying these libraries into a single organizational group. Harvard's digital library activities are designed to provide a common technology infrastructure, consulting assistance, and guidance on policy issues to all of the Harvard libraries. Unlike many universities, Harvard has established major funding to support the development of a comprehensive infrastructure for digital libraries.

The $12 million, 5-year program is called the "Library Digital Initiative," deliberately emphasizing the library orientation of the program. The goal of centrally funding this initiative to serve the decentralized network of libraries is to create incentive for participation in a common, standardized solution featuring a robust production. It environment and common practices for digitizing, reformatting, metadata creation, digital licensing, preservation, and migration of objects, etc.

According to Harvard Library website (https://library.harvard.edu/visit-about/about-harvard-library) Harvard Library Network is the champions of curiosity. With the aim to be global leaders in expanding world knowledge and intellectual exploration. The library is engaged with communities in the creation and sharing of new knowledge, connecting them with the vast collections that they curate and steward through collaborations around the world. At its core, the Harvard Library mission for nearly four centuries has been to advance the learning, research, and pursuit of truth that are at the heart of Harvard. Library efforts are motivated and powered by working collaboratively, embracing diverse perspectives, championing access, aiming for the extraordinary, and always leading with curiosity.

The Harvard Library Network have over 800 staff engaging with scholars and students across more than 25 libraries and around the world. With over 20 million physical and digital items, its vast collections are renowned for their global reach and depth, with resources spanning the development of all disciplines and more than 460 languages. The rare and special collections are amongst the most remarkable in the world, ranging from medieval manuscripts to sound recordings of modern poets and from early maps to digital images. They collect collaboratively with peer institutions and facilitate international open access, seeking to build a diverse, open corpus of knowledge with scholars and citizens everywhere. Established in 1638, the oldest library system in the United States and the world's largest academic library.

1.8.4 JSTOR Digital Library

JSTOR began as a pilot project sponsored by the Andrew W. Mellon Foundation in 1990. Ten journals in economics and history were selected in initially, and all back issues were digitized. High resolution images of each page were captured and linked to a text file generated with optical character recognition software. (www.jstor.org) Table of contents indices were developed as well, to provide search and retrieval of journal content of interest. When five test libraries showed enthusiasm about the space savings and ease of access, JSTOR was born.

In 1995 JSTOR was established as an independent non-profit organization with the goals of providing a trustworthy archive of important scholarly journals, reducing the cost of accessing these materials, and assuring their long-term preservation. JSTOR now provides more than 1,200 user institutions worldwide access to JSTOR

provides access to more than 12 million journal articles, books, images, and primary sources in 75 disciplines.

It helps the users to explore a wide range of scholarly content through a powerful research and teaching platform. They collaborate with the academic community to help libraries connect students and faculty to vital content while lowering costs and increasing shelf space, provide independent researchers with free and low-cost access to scholarship, and help publishers reach new audiences and preserve their content for future generations.

JSTOR is part of ITHAKA, a not-for-profit organization that also includes Artstor, Ithaka S+R, and Portico.

The entire JSTOR archive requires more than 2.2 terabytes of storage for more than fifteen million pages of journal content. Usage has been growing at a rate of 50% per year, with rapid growth among international users. There is also increasing emphasis on allowing access to the JSTOR archive by secondary schools and public libraries.

1.9 Digital Libraries Online E-Journals and E-Books

Today, many large publishers around the world, both for-profit and not-for-profit, maintain Web sites that make available their full collections of print journals (with only limited back-file runs, so far) to subscribers. Given this penetration of new technologies into scholarly, scientific and popular journal and magazine publication, a list or directory of print journals will one day soon no longer be needed. In this decade (or even half-decade), most print magazines and journals will have a Web version, if they do not already have one.

Furthermore, although for the moment it is convenient to think of print and Web versions as providing the same or identical information, the two styles are already beginning to pull apart and will only diverge further. Not only will the same name ultimately denote collections of content that are in fact very different, but some of the e-journals also will evolve into new genres.

The prevailing vision is that this wealth of journal literature will be linked through indexing services and this vision is rapidly being realized. For as long as print versions continue to be published, tracking and collecting will be much more difficult. The most significant problem for ubiquitous electronic access is that of long-term achievability and preservation, a not to maintain costly parallel print and electronic systems.

So far, very few electronic journals (or any other resource) have had to survive on the internet for even one decade. While some experts say that long-term

sustainability is a trivial matter. Studies suggest high costs and tremendous uncertainties.

Fundamental to these uncertainties is matter of ownership, which libraries rarely have, given that electronic information produces no fixed artifact for characterizes scholarly and research journals, the committee found similar or even faster growth in the even broader universe of all continuing publications that includes popular magazines, series and newspapers as well as annual reports, directories, series, and so on. Standard periodical indexing and abstracting services began to become available electronically in the 1970s through specialized vendors such as Dialog and BRS, whose proprietary systems required the mediation of expert easy-to-use Web interfaces for any licensed subscriber.

Thus, it is no surprise that the journal articles cited by these sources would become quickly available online. However, books-such as novels and scholarly monographs, for example-have seemed far less susceptible to electronic transformation. While some books are consulted in bits and bytes (for particular facts or small sections), many (the argument goes) need to be deliciously savored and contemplated from beginning to end, and an online screen is hostile to such prolonged congenial or intense reading.

The vision now being articulated by many players in the book publishing industry and its partners(in printing, distribution, and software) is that the full text of all published books, at least from mainstream publishers, will exist on vast electronic information servers, there to be channeled to the output of the reader's choice: traditional print formats or digital formats (by accessing a local copy on a PC or portable device or by viewing a remote copy through the Web).that is, the authoritative source file of many books may soon be an electronic version that can generate various derivative versions.

The e-book may be on the verge of acceptance and success because of the convergence of large computer servers, big network pipelines, rapid progress in developing e-book standards, rapid progress in developing e-book standards, and the increasing sophistication and utility of handheld book-reading devices, as well as business partnerships to take advantage of all this.

Like e-journals, e-books also have a history that goes back to the 1980s.
In the mid-1990s, an entrepreneur making a car trip with his wife had a vision of "the world's largest bookstore," a virtual shop in which the discerning reader could obtain every book-or nearly every book-currently in print. Jeff Bezos, founder and CEO of amazon.com, Time Magazine's person of the year for 1999, achieved his dream: situating bookselling at the core of e-commerce and radicalizing the bookstore concept, to say nothing of notions of business success.

Through the Web, the Amazon.com shopper fills a virtual shopping cart with the desired books, which are delivered a day or two or three later to the reader's address of choice. Through Amazon.com, large collections of books have come one step nearer to their readers and no library has played a part in this dramatic convergence of reader and book. In the case of the virtual superstore, the reader controls the atmosphere, which need not be as public as a library or bookshop and may be as comfortable as reading in a lounge chair wearing fuzzy slippers and a dressing gown. Most significantly, the rise of the book superstore has implicitly changed the overall economics of access to books and information.

Where once a good public library was the best and most accessible source of materials for many, if not most, communities, bookstores of similar size may be a few doors down the block, open longer hours, and with enough copes of popular titles to satisfy almost all comers.

And, most libraries and physical bookstores are dwarfed by online bookstores. These book superstores offer a remarkably wide range of library-like services lectures, discussion groups, ready access to books, a sense of community. What they do not yet offer is information service-labor intensive; predict that they will not continue to expand their range.

For now, it is most noticeably the information navigation functions of the library-the instruction in finding, filtering, and evaluating information from a welter of available sources-and in provision of historical depth that the bookstores make the least attempt to supplant. Borders began its career with claims about the literacy and helpfulness of its staff, all of whom had to pass a test to demonstrate their book lore as a condition of employment despite this, the library is less than ever a primary supplier of access to new and current books.

Libraries continue however, to be strong in providing access to both older and more specialized material. Another implication of the shift in the economy of information is that collecting and storing materials are arguably somewhat less the jobs of the library than before. The current emphasis is on services. Indeed, many public libraries have always had this bias, retaining from among older book only those of continuing interest to their readers, while clearing out shelf space for what a current generation demands.

In any case, the superstores and Amazon.com have by no means reached the end state of the publishing industry's book-to-reader vision, nor do libraries seem to have much of a role in that next vision either.

The concept of books on demand has been in gestation for some time. For example, in the early 1990s, Xerox partnered with a few large publishing houses in an OD experiment. The products were acceptable, but computer servers and network

pipelines were less capacious that today and so the results were slow. The need to resolve right and permissions issues also posed a significant challenge. The concept needed time to ripen, and ripen it has.

Most individuals, organizations and publishers agree that at the very least the traditional book format is facing competition from formats that do some things better. In 1999, the American Historical Association announced an electronic book prize to be awarded annually for several years to half a dozen brilliant dissertations in various fields of history. Dissertations that take full advantage of the converging can offer and books cannot. The results will enliven the books, attract readers, promote the new medium as a viable one for serious scholarship and give young scholars a leg up.

1.10 Digital library and Scholarly Communication

The multitude of electronic databases, the rapid growth of Web sites, the increase in the number of electronically available print journals, and the availability of numerous full-text resources such as reports, dissertations, and electronic books all represent a dramatic change in the dissemination of scholarly and cultural content. The history of this revolution is short and there is still much to learn, on all levels and in all areas. Not a great deal has been published about how the great libraries are transforming themselves to greet the electronic age, although Maurice B Line noted that three factors have caused an almost ceaseless questioning of the roles and futures of national libraries:

1. Automation, information, and communications technology;

2. The intrusion of the private sector into areas formerly sacred to libraries; and

3. The globalization of libraries.

Libraries have accordingly begun to move aggressively in the direction of becoming hybrid libraries (i.e., libraries that embrace information in numerous formats, now including electronic formats).

The transformation of library cultures and practice by adoption of information technology continues apace. An increasing emphasis on service and a decreasing emphasis on collection have already been noted. Libraries are incorporating electronic technologies and services into the everyday work of all staff by doing a number of things:

- Working for the broadest possible access for readers in the electronic environment. Not only are libraries seeking technological standards' and

presentation of resources in forms accessible to the broadest range of readers, but they are also lobbying to advance the public policy debate in ways that support broad access for the good of society as a whole.

- Reallocating an increasing and visible portion of collections budgets to the electronic resources needed by their readers.
- Building collections of digital resources that, while not yet rivalling traditional collections in scope and bulk, are substantial, of high value, and integrated in the traditional patterns of collection and use.
- Working to shape and support initiatives such as community education, online course support, Web page design, teaching specialist electronic resources, and digitizing of materials for these programs-all with a view to making educational opportunities as broad, rich, and accessible as possible. Lifelong learning is the opportunity and the goal, and "distance learning" us the current buzzword for the tactics librarians seek to support.
- Finding new ways to measure the usage patterns and behaviors of readers, so as to anticipate and support their needs, bringing the right resources into play for readers. The digital environment facilitates such measurement and, accordingly, such feedback, giving a better allocation of resources than has ever been possible with print media.
- Devoting increasing effort to more sophisticated reader services associated with single and multiple electronic resources. Librarians are more often than ever teachers of how to use electronic resources, and readers spend less time pursuing simple factual information at traditional reference desks.
- Cooperating with other libraries in setting up networks that make libraries effectively a single virtual (through the locator tool of interoperable online catalogs) institution that can deliver physical materials, via advanced interlibrary loans and document delivery, to more and more readers more effectively -and more cost-effectively -than ever.
- Delivering physical materials by electronic means. As physical materials become increasingly deliverable at a distance, libraries are putting more and more electronic delivery services into operation.
- Partnering with other participants in the creations and dissemination of knowledge. Libraries can for example, work with individual authors, organizations, publishers (commercial and non-commercial), booksellers, and software companies to create and make available functional and well used online resources.
- Digitizing and making available to readers materials already in library collections and special collections. Such materials would include, in particular out-of-copyright material, image collections, sheet music, maps, and other traditional library treasures.

- Subscribing to online services that provide statistical data. Libraries would help readers learn to manipulate services containing anything from historical census data to financial market data.
- Creating multimedia servers for music, film, and other media. At the same time, thorny questions of access and permitted use must be addressed, and the technological capability to handle significant quantities of such material must be developed.
- Using the new generation of library management systems as a springboard not only for integrating forms of access to a wide range of materials and formats but also for reengineering the entire workflow and back-office processes of traditional librarianship. The technical services of libraries are becoming increasingly business-like, streamlined, and closely managed, with closer links than ever to vendors through electronic data interchange (EDI) and other forms of electronic interaction that work to the advantage of all parties.
- Working to understand the technical demands, possibilities, and long-term costs and responsibilities of digital media as instruments for the preservation of library information, including material from traditional print media (e.g., the contents of books printed on acid-based paper) and material created in digital form. When we fully understand the challenges of moving digitally preserved information from format to format, from one hardware and software systems to a new hardware and software system, we will have made great progress in solving what many think is the biggest remaining problem in establishing truly functional and satisfactory digital libraries.
- Working through the issues that must be faced in deciding which kinds of resources are best maintained locally, library by library, and which resources are best maintained elsewhere, whether by publishers, vendors, library consortia, or third parties. Traditional librarianship achieves security and preservation by having redundant physical copies: the challenge now is to balance redundancy (and thus security) with optimal efficiency and to avoid unnecessary duplication of effort.
- Understanding evolving legal regimes such as copyright and licensing. In this arena, librarians seek not only to understand but also to shape and influence developments, thus securing agreements that offer reader's high-quality, reliable, and permanent access to resources.
- Exercising responsible stewardship of library resources, which are usually purchased with public funds or from not for-profit institutional budgets. Such stewardship requires keen understanding of the business models and economics of the new information sources in an environment in which libraries find themselves increasingly offered not ownership but access not

a once-for-all price but something closer to annual subscription or by-the drink pricing.

- Cultivating an expertise in technology matters. The technological infrastructure of a library now faces a new degree of volatility and continuing costs as equipment and software need upgrading. The marketplaces makes it literally impossible to choose not to play the upgrade game: in a very short time, a library's information would simply become unavailable if it persisted in using even slightly outmoded operating systems or software.
- Continuously upgrading human resources and skills. The librarians and support staff at this time of transformation must undergo no less arduous a series of "upgrades." As in other sectors of our economy, it is impossible in the library sector for staff to acquire and practice skills and then use them for a lifetime; instead, they must grow and adapt, and there are real and substantial costs for supporting the necessary training and for paying a more highly skilled staff.
- Seeking new funding sources and opportunities. Traditional funding sources-annual budgets doled out by the government or not-for -profit organization with a tiny annual increase-no longer suffice.

Librarians are increasingly engaged in entrepreneurial efforts, whether soliciting research and development funding from granting agencies, developing partnerships with other entities in the library sector, or participating in cost-recovery projects with the commercial sector that serves and interacts with the library community.

1.11 Conclusion

In broad sense digital library is a computerized system that allows users to obtain a coherent means of access to an organized, electronically stored repository of information and data. The digital library is on electronic library consisting of information in the digital, analog and digitized form. Librarians tend to speak for a broader definition of the term "Library". They see a library as an organisation that secures the selection, conservation, organisation, preservation and the access to information that is vital for the members of the specific organisation.

Current information needs are being provided in electronic form with varying success in public, college and research libraries around the world.

Research libraries have only begun to take on the provision, organization and preservation of information with the same long-term commitment they have made for print materials. It is an expensive, uncharted and difficult task. But until the

long-term commitments are undertaken, many currently proposed solutions will have only temporary effects. The ability of the scholarly community to give serious weight to electronic information depends upon their trust in such information being dependably available, with authenticity and integrity maintained.

Changes in scholarly publishing to help alleviate the serials crisis, for example, are usually thought to be bound up with the prestige of electronic journals in the academic tenure process. The ability of the academy to count on long-term, secure existence of electronic scholarly work will be an important determinant of the success of academic electronic publishing. Libraries and universities have a stake in helping electronic publishing to succeed, and therefore have an interest in establishing secure digital research libraries. Users' needs will continue to be what they long have been.

The locus of information may be called the electronic storage repository. Over time, we will learn how collection development plays out in an access environment as well as in an ownership environment. It is sometimes loosely proposed that libraries need not acquire electronic information, for it will be available somewhere on the network.

Such proposals ignore the obvious truth that some institution must still, in the end, take responsibility for the information. That has always been a definition of the library responsibility. There will be many electronic storage repositories, responding both to requirements of redundancy and to the individual needs of institutions.

In contrast to print collections, it is unlikely that there will be a high degree of content duplication across many electronic repositories, since for most purposes existence in a single place allows world-wide access.

Self-Assessment Questions

1. Define digital library and discuss its components, features, and characteristics.
2. What is digital imaging discuss various forms of digital imaging and advantages of digital contents?
3. Describe various online digital library services provided by scholarly publishing vendors.
4. What is meant for scholarly communication? Describe the role of digital library in the provision of scholarly communication.

Activity:

1. Visit the online portal of ProQuest online digital library and describe its various services provided to researchers and librarians.

Recommended Readings:

1. Arms, W. Y. (2000). *Digital Libraries*. Cambridge, MA: The MIT Press.

2. Baker, D., & Evans, W. (2008). *Digital Library Economics: Digital library economics: An academic perspective*. Oxford, UK: Chandos Publishing.

3. Borgman, C. (2000). *From Gutenberg to the Global Information Infrastructure: Access to information in the networked world*. Cambridge, MA: MIT Press.

4. Cornell University Library. (n.d.). Retrieved from http://www.library.cornell.edu/

5. Greenstein, D. (2000). Digital Libraries and Their Challenges. *Library Trends*, *49*(Fall), 290–303.

6. Greenstein, D. (2002, March 15). Digital Library Orchestrating Digital Worlds. *Library Journal.*

7. Hedstrom, M. (2004). Digital preservation: A time bomb for digital libraries. *Computers and the Humanities*, *31*(3), 189–202. doi:10.1023/A:1000676723815

UNIT-2

DIGITIZATION: METHODS, TOOLS, TECHNOLOGIES AND TECHNIQUES

Complied by: Muhammad Jawwad

Reviewed by: Dr. Pervaiz Ahmed

CONTENT

INTRODUCTION

This unit is designed to inoculate in students the digitization process, key components, features of digitization and characteristics of digital library development. Students can establish a background knowledge of digital imaging, criterion of selecting materials for digitization, and challenges faced by libraries in digital preservation. It will also describe the role of electronic format in provision of scholarly communications, process of digitization and tools required for digitization.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is Digitization?
2. Why digitization is necessary?
3. What is the criterion of selecting materials for digitization and its various factors?
4. What steps are involved in the process of digitization?
5. What tools required for digitization and how digital images are organized?

2.1 Introduction

All recorded information in a traditional library is analogue in nature. The analogue information can include printed books, periodical articles, manuscripts, cards, photographs, vinyl disks, video and audio tapes. However, when analogue information is fed into a computer, it is broken down into 0s and 1s changing its characteristics from analogue to digital. These bits of data can be re-combined for manipulation and compressed for storage. Voluminous encyclopaedias that take-up yards of shelf-space in analogue form can fit into a small space on a computer drive or stored on to a CD ROM disc, which can be searched, retrieved, manipulated and sent over the network. One of the most important traits of digital information is that it is not fixed in the way that texts printed on a paper are. Digital texts are neither final nor finite and are not fixed either in essence or in form except, when it is printed out as a hard copy.

Flexibility is one of the chief assets of digital information. An endless number of identical copies can be created from a digital file, because a digital file does not decay by copying. Moreover, digital information can be made accessible from remote location simultaneously by a large number of users. Digitisation is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) into digital format. In most library applications, digitisation normally results in documents that are accessible from the web site of a library and thus, on the Internet. Optical scanners and digital cameras are used to digitise images by translating them into bit maps. It is also possible to digitise sound, video, graphics, animations, etc.

Digitisation is not an end in itself. It is the process that creates a digital image from an analogue image. Selection criteria, particularly those which reflect user needs are of paramount importance. Therefore, the principles that are applicable in traditional collections development are applicable when materials are being selected for digitisation. However, there are several other considerations related to technical, legal, policy, and resources that become important in a digitisation project. Digitisation is one of the three important methods of building digitised collections. The other two methods include providing access to electronic resources (whether free or licensed) and creating library portals for important Internet resources.

2.2 Needs of Digitization

Digitising a document in print or other physical media (e.g., sound recordings) makes the document more useful as well as more accessible. It is possible for a user

to conduct a full-text search on a document that is digitised and OCRed. It is possible to create hyperlinks to lead a reader to related items within the text itself as well as to external resources. Ultimately, digitisation does not mean replacing the traditional library collections and services; rather, it serves to enhance them. A document can be converted into digital format depending on the objective of digitisation, end user, availability of finances, etc. While the objectives of digitisation initiatives differ from organization to organization, the primary objective is to improve the access. Other objectives include cost savings, preservation, keeping pace with technology and information sharing. The most significant challenges in planning and execution of a digitisation project relate to technical limitations, budgetary constraints, copyright considerations, lack of policy guidelines and lastly, the selection of materials for digitisation.

While new and emerging technologies allow digital information to be presented in innovative ways, the majority of potential users are unlikely to have access to sophisticated hardware and software. Sharing of information among various institutions is often restricted by the use of incompatible software. One of the main benefits of digitisation is to preserve rare and fragile objects by enhancing their access to multiple numbers of users simultaneously. Very often, when an object is rare and precious, access is only allowed for a certain category of people. Going digital could allow more users to enjoy the benefit of access. Although, digitisation offers great advantages for access like, allowing users to find, retrieve, study and manipulate material, it cannot be considered as a good alternate for preservation because of ever changing formats, protocols and software used for creating digital objects.

There are several reasons for libraries to go for digitisation and there are as many ways to create the digitised images, depending on the needs and uses. The prime reason for the digitisation is the need of the user for convenient access to high quality information. Other important considerations are:

2.3 Quality of Digitization and Preservation

The digital information has potential for qualitative preservation of information. The preservation quality images can be scanned at high resolution and bit depth for best possible quality. The quality remains the same inspite of multiple usage by several users. However, caution need to be exercised while choosing digitised information as preservation media.

1. Multiple Referencing: Digital information can be used simultaneously by several users at a time.

2. Wide Area Usage: Digital information can be made accessible to distant users through the computer networks over the Internet.

3. Archival Storage: Digitisation is used for restoration of rare material. The rare books, images or archival material are kept in digitised format as a common practice.

4. Security Measure: Valuable documents and records are scanned and kept in digital format for safety.

2.4 Criteria of Selecting Material for Digitization

To begin the process of digitisation, first of all, we need to select documents for digitisation. The proccess of selection of material for digitisation invloves identification, selection and prioritization of documents that are to be digitized. If an organization generates contents, strategies may be adopted to capture data that is "borne digital". If documents are available in digital form, it can be easly converted into other formats. If the selected material is from the external sources, IPR issues need to be resolved. If material being digitised is not available in public-domain then it is important to obtain permission from the publishers and data suppliers for digitisation.

Moreover, decision may be taken whether to OCR the digitized images. Documents selected for digitisation may already be available in digital format. It is always economical to buy e-media, if available, than their conversion. Moreover, over-sized material, deteriorating collections, bound volumes of journals, manuscripts, etc. would require highly specialized equipment and highly skilled manpower.

The documents to be digitised may include text, line art, photographs, colour images, etc. The selection of document need to be reviewed very carefully considering all the factors of utility, quality, security and cost. Rare and much in demand documents and images are selected as first priority without considering the quality.

Factors that may be considered before selecting different media for digitisation include:

- **Audio:** The sound quality has to be checked and require corrections made together by the subject expert and computer sound editor.

- **Video:** The video clippings are normally edited on Beta max tapes which can be used for transferring it on digital format. While editing colour tone, resolution is checked and corrected.

- **Photographs:** The selection of photographs is very crucial process. High resolution is required for photographic images and slides. Especially the quality, future need and the copyright aspects have to be checked.

- **Documents:** Documents which are much in demand, too fragile to handle, and rare in availability are reviewed and selected for the process. If the correction of literary value demands much input, then documents are considered for publication rather than digitisation.

2.5 Steps in the Process of Digitization

The following four steps are involved in the process of digitisation. Software, variably called document image processing (DIP), Electronic Filing System (EFS) and Document Management Systems (DMS) provides all or most of these functions:

2.5.1 Scanning

Electronic scanners are used for acquisition of an electronic image into a computer through its original that may be a photograph, text, manuscript, etc. An image is "read" or scanned at a predefined resolution and dynamic range. The resulting file, called "bit-map page image" is formatted (image formats described elsewhere) and tagged for storage and subsequent retrieval by the software package used for scanning. Acquisition of image through fax card, electronic camera or other imaging devices is also feasible. However, image scanners are most important and most commonly used component of an imaging system for the transfer of normal paper-based documents.

Steps in the Process of Scanning using a Flatbed Scanner

Step 1. Place picture on the scanner's glass
Step 2. Start scanner software
Step 3. Select the area to be scanned
Step 4. Choose the image type
Step 5. Sharpen the image
Step 6. Set the image size
Step 7. Save the scanned image using a desirable format (GIF or JPEG)

2.5.2 Indexing

If converting a document into an image or text file is considered as the first step in the process of imaging, indexing these files comprises the second step. The process of indexing scanned image involves linking of database of scanned image to a text database. Scanned images are just like a set of pictures that need to be related to a text database describing them and their contents.

An imaging system typically stores a large amount of unstructured data in a two file system for storing and retrieving scanned images. The first is traditional file that has a text description of the image (keywords or descriptors) along with a key to a second file. The second file contains the document location. The user selects a record from the first file using a search algorithm. Once the user selects a record, the application keys into the location index finds the document and displays it.

Most of the document imaging software packages through their menu drive or command driven interface, facilitate elaborate indexing of documents. While some document management system facilitate selection of indexing terms from the image file, others allow only manual keying in of indexing terms. Further, many DMS packages provides OCRed capabilities for transforming the images into standard ASCII files. The OCRed text then serve as a database for full-text search of the stored images.

2.5.3 Store

The most tenacious problem of a document image relates to its file size and, therefore, to its storage. Every part of an electronic page image is saved regardless of the presence or absence of ink. The file size varies directly with scanning resolution, the size of the area being digitized, and the style of graphic file format used to save the image. The scanned images, therefore, need to be transferred from the hard disc of scanning workstation to an external large capacity storage device such as an optical disc, CD ROM / DVD ROM disc, snap servers, etc. While the smaller document imaging system may use offline media, which need to be reloaded when required, or fixed hard disc drives allocated for image storage. Larger document management system use auto-changers such as optical jukeboxes and tape library systems. The storage required by the scanned image varies and depends upon factors such as scanning resolution, page size, compression ratio and page content.

Further, the image storage device may be either remote or local to the retrieval workstation depending upon the imaging systems and document management system used.

2.5.4 Retrieve

Once scanned images and OCRed text documents have been saved as a file, a database is needed for selective retrieval of data contained in one or more fields within each record in the database. Typically, a document imaging system uses at least two files to store and retrieve documents. The first is traditional file that has a text description of the image along with a key to the second file. The second file contains the document location. The user selects a record from the first-file using a search algorithm. Once the user selects a record, the application keys into the location index finds the document and displays it. Most of the document management system provides elaborate search possibilities including use of Boolean and proximity operators (and, or, not) and wild cards. Users are also allowed to refine their search strategy. Once the required images have been identified their associated document image can quickly be retrieved from the image storage device for display or printed output.

2.6 Digitization: Input and Output Options

A document can be converted into digital format depending on the objective of digitisation, end user, availability of finances, etc. There are four basic approaches that can be adapted to translate from print to digital:
Scanned as Image Only OCR and Retaining Page Layout Retaining Page Layout using Acrobat Capture; and Re-keying the Data

2.6.1 Scanned as Image Only

Image only is the lowest cost option in which each page is an exact replica of the original source document. Several digital library projects are concerned
by providing digital access to materials that already exists with traditional libraries in printed media. Scanned page images are practically the only reasonable solution for institutions such as libraries for converting existing paper collection (legacy documents) without having access to the original data in computer processible formats convertible into HTML / SGML or in any other structured or unstructured text. Scanned page images are natural choice for large-scale conversions for major digital library initiatives. Printed text, pictures and figures are transformed into computer-accessible forms using a digital scanner or a digital camera in a process called document imaging or scanning.

The digitally scanned images are stored in a file as a bit-mapped page image,
irrespective of the fact that a scanned page contains a photograph, a line drawing or text. A bit-mapped page image is a type of computer graphic, literally an electronic picture of the page which can most easily be equated to a facsimile image of the page and as such they can be read by humans, but not by the computers,

under stably “text” in a page image is not searchable on a computer using the present-day technology. An image-based implementation requires a large space for data storage and transmission.

Capturing page image format is comparatively easy and inexpensive, therefore, it is a faithful reproduction of its original maintaining page integrity and originality. The scanned textual images, however, are not searchable unless it is OCRed, which in itself, is highly error prone process specially when it involves scientific texts. Options of technology for converting print to digital are given separately.

Since OCR is not carried out, the document is not searchable. Most scanning software generate TIFF format by default, which, can be converted into PDF using a number of software tools. Scan to TIFF / PDF format is recommended only when the requirement of project is to make documents portable and accessible from any computing platform. The image can be browsed through a table of contents file composed in HTML that provides link to scanned image objects.

2.6.2 Optical Character Recognition (OCR) and Retaining Page Layout

The latest versions of both Xerox’s Text Bridge and Caere’s Omnipage incorporate technology that allow the option of maintaining text and graphics
in their original layout as well as plain ASCII and word-processing formats. Output can also include HTML with attributes like bold, underline, and italic which are retained. Retaining Layout after OCR A scanned document is nothing more than a picture of a printed page. It can not be edited or manipulated or managed based on their contents. In other words, scanned documents have to be referred to by their labels rather than characters in the documents.

OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing file. OCR or text recognition is the process of electronically identifying text in a bit-mapped page image or set of images and generate a file containing that text in ASCII code or in a specified word processing format leaving the image intact in the process.

2.6.3 Retaining Page Layout Using Acrobat Capture

The Acrobat Capture latest version 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, so that the scanned and OCRed copy never over- or under-shoots the page. Accordingly, it treats unrecognizable text as images that are pasted in its place. Such images are perfectly readable by anyone by looking at the PDF file, but will be absent from the editable and searchable text file. In contrast, ordinary OCR programs treat unrecognized text as tilde or some other special character in the ASCII

output. Acrobat Capture can be used to scan pages as images, image +text and as normal PDF, all the three options retain page layout.

2.6.4 Re-Keying

A classic solution of this kind would comprise of keying-in the data and its verification. This involves a complete keying of the text, followed by a full rekeying by a different operator, the two keying-in operation might take place simultaneously. The two keyed files are compared and any errors or inconsistencies are corrected. This would guarantee at least 99.9% accuracy, but to reach 99.955% accuracy level, it would normally require full proof-reading of the keyed files, plus table lookups and dictionary spell checks.

2.7 Technology used in Digitization

Digital images, also called "bit-mapped page image" are "electronic photographs" composed or set of bits or pixels (picture elements) represented
by "0" and "1". A bit-mapped page image is a true representation of its original in terms of typefaces, illustrations, layout and presentation of scanned documents. As such information or contents of "bit-mapped page image" cannot be searched or manipulated unlike text file documents (or ASCII).

However, an ASCII file can be generated from a bit-mapped page image using an optical character recognition (OCR) software such as Xerox's
Text Bridge and Caere's OmniPage. The quality of digital image can be monitored at the time of capture by the following factors:

- Bit depth / dynamic range
- Resolution
- Threshold
- Image enhancement

Terminology associated with technological aspects of digitisation described below is given in the keywords. Students are advised to understand the terminology, specially bit, byte and pixel before going through the unit.

2.7.1 Bit Depth or Dynamic Range

The number of bits used to define each pixel determines the bit depth. The greater the bit depth, the greater the number of gray scale or colour tones that can be represented. Dynamic range is the term used to express the full range of total variations, as measured by a densitometer between the lightest and the darkest of a document.

Table 1. No. of bits used for representing shades in colour and gray-scale scanning

Sl. No.	No. of Bits	No. of bits / shades	No. of shades	No. of Shades / pixel
1	2	2	$2^2=4$	$4^3=64$
2	4	3	$2^3=8$	$8^3=512$
3	8	4	$2^4=16$	$16^3=4096$
4	16	5	$2^5=32$	$32^3=32,768$
5	32	6	$2^6=64$	$64^3=262144$
7	64	7	$2^7=128$	$128^3=2097152$
8	128	8	$2^8=256$	$256^3=16777216$

Digital images can be captured at varied density or bits pixel depending upon i) the nature of source material or document to be scanned; ii) target audience or users; and iii) capabilities of the display and print subsystem that are to be used. Bitonal or black & white or binary scanning is generally employed in libraries to scan pages containing text or the drawings.

Bitonal or binary scanning represents one bit per pixel (either “0” (black) or “1” (white). Gray scale scanning is used for reliable reproduction of intermediate or continuous tones found in black & white photographs to represent shades of grey.

Multiple numbers of bits ranging from 2-8 are assigned to each pixel to represent shades of grey in this process. Although each bit is either black or
white, as in the case of bitonal images, but bits are combined to produce a level of grey in the pixel that is, black, white or somewhere in between.

Lastly colour scanning can be employed to scan colour photographs. As in the case of grey-scale scanning, multiple bits per pixels typically 2 (lowest quality) to 8 (highest quality) per primary colour are used for representing colour. Colour images are evidently more complex than grey scale images, because it involves encoding of shades of each of the three primary colours, i.e. red, green and blue (RGB). If a coloured image is captured at 2 bits per primary colour, each primary colour can have 22 or 4 shades and each pixel can have 43 shades for each of the three primary colours. Evidently, increase in bit depth increases the quality of image captured and the space required to store the resultant image. Generally speaking, 12 bits per pixel (4 bits per primary colour) is considered. minimum pixel depth for good quality colour image. Most of today’s colour scanner can scan at 24-bit colour (8 bit per primary colour).

2.7.2 Resolution

The resolution of an image is defined in terms of number of pixel (picture elements) in a given area. It is measured in terms of dot per inch (dpi) in case of an image file and as ratio of number of pixel on horizontal line x Number of pixel in vertical lines

in case of display resolution on a monitor. Higher the dpi is set on the scanner, the better the resolution and quality of image and larger the image file.

Regardless of the resolution, image quality of an image can be improved by capturing an image in greyscale. The additional gray-scale data can be processed electronically to sharpen edges, file-in characters, remove extraneous dirt, remove unwanted page strains or discoloration, so as to create a much higher quality image than possible with binary scanning alone. A major drawback in gray scale is that there is large amount of data capture. It may be noted that continuing increase in resolution will not result in any appreciable gain in image quality after some time, except for increase in file size. It is thus important to determine the point where sufficient resolution has been used to capture all significant detail present in the source document. The black and white or bitonal images (textual) are scanned most commonly at 300 dpi that preserve 99.9% of the information contents of a page and can be considered as adequate access resolution. Some preservation projects scans at 600 dpi for better quality. A standard SVGA/VGA monitor has a resolution of 640 x 480 lines while the ultra-high monitors have a resolution of about 2048 x 1664 (about 150 dpi).

2.7.3 Threshold

The threshold setting in bitonal scanning defines the point on a scale, usually ranging from 0 - 255, at which grey values will be interpreted as black or white pixels. In bitonal scanning, resolution and threshold are the key determinants of image quality. Bitonal scanning is best suited to high-contrast documents, such as text and line drawings. Gray scale or colour scanning is required for continuous tone or low contrast for documents such as photographs. In grey scale/colour scanning both resolution and bit depth combine to play significant roles in image quality. In Line art mode, every pixel has only two possible values. Every pixel will either be black or white.

The Line art Threshold control determines the decision point about brightness determining if the sampled value will be a black dot or a white dot. The normal Threshold default is 128 (the midrange of the 8-bit 0 - 255 range). Image intensity values above the threshold are white pixels, and values below the threshold are black pixels. Adjusting threshold is like a brightness setting to determine what is black and what is white. Threshold for text printed on a coloured background or cheap-quality paper like newsprint has to be kept at lower range. Reducing threshold from128 to about 85 would greatly improve the quality of scan. Such adjustments would also improve the performance of OCR software.

2.7.4 Image Enhancement

Image enhancement process can be used to improve scanned images at a cost of image authenticity and fidelity. The process of image enhancement is, however, time consuming, it requires special skills and would invariably increase the cost of conversion. Typical image enhancement features available in a scanning or image editing software include filters, tonal reproduction, curves and colour management, touch, crop, image sharpening, contrast, transparent background, etc. In a page scanned in grey-scale, the text /line art and half tone areas can be decomposed and each area of the page can be filtered separately to maximize its quality. The text area on page can be treated with edge sharpening filters, so as to clearly define the character edges, a second filter could be used to remove the high-frequency noise and finally another filter could fill-in broken characters. Grey-scale area of the page could be processed with different filters to maximize the quality of the halftone.

2.8 Digitization and Compression Process

Image files are evidently larger than textual ASCII files. It is thus necessary to compress image files so as to achieve economic storage, processing and transmission over the network. A black & white image of a page of text scanned at 300 dpi is about 1 mb in size where as a text file containing the same information is about 2-3 kb. Image compression is the process of reducing size of an image by abbreviating the repetitive information such as one or more rows of white bits to a single code. The compression algorithms may be grouped into the following two categories:

2.8.1 Lossless Compression

The conversion process converts repeated information as a mathematical algorithm that can be decompressed without loosing any details into the original image with absolute fidelity. No information is “lost” or “sacrificed” in the process of compression. Lossless compression is primarily used in bitonal images.

2.8.2 Lossy Compression

Lossy compression process discards or minimize details that are least significant or which may not make appreciable effect on the quality of image. This kind of compression is called “lossy” because when the image that is compressed using “Lossy” compression techniques is decompressed, it will not be an exact replica of the original image. Lossy compression is used with grey-scale / colour scanning.

Compression is a necessity in digital imaging but more important is the ability to output or produce the uncompressed true replica of images. This is especially important when images are transferred from one platform to another or are handled by software packages under different operating system. Uncompressed images

often work better than compressed images for different reasons. It is thus suggested that scanned images should be either stored as uncompressed images or at the most as lossless compressed images. Further, it is optimal to use one of the standard and widely supported compression protocols than a proprietary one, even if it offers efficient compression and better quality. Attributes of original documents may also be considered while selecting compression techniques. For example ITU G-4 is designed to compress text where as JPEG, GIF and Image PAC are designed to compress pictures. It is important to ensure migration of images from one platform to another and from one hardware media to another. It may be noted that highly compressed files are more prone to corruption than uncompressed files.

2.9 Digitization: File Formats and Media Types

A defined arrangement for discrete sets of data that allow a computer and software to interpret the data is called a file format. Different file formats are used to store different media types like text, images, graphics, pictures, musical works, computer programs, databases, models and designs video programs and compound works combining many types of information.

Although almost every type of information can be represented in digital form, a few important file formats for text and images typically applicable to a library-based digital library are described here. However, every object in a digital library needs to have a name or identifier which distinctly identifies its type and format. This is achieved by assigning file extensions to the digital objects. The file extensions in a digital library typically denote formats, protocols and right management that are appropriate for the type of material.

Text and image-based contents of a digital library can be stored and presented as i) simple text or ASCII (American Standard Code for Information Interchange; ii) unstructured text; iii) Structured text (SGML or HTML or XML); iv) page description language and v) page image formats. Textual and image formats are dealt elaborately in the section on “Technical Infrastructure of Digital Library”.

2.10 Tools of Digitization

Digital imaging is an inter-linked system of hardware, software, image database and access sub-system with each having their own components. Tools used for digitisation include several core and peripheral systems. An image scanning system may consist of a stand-alone workstation where most or all the work is done on the same workstation or as a part of a network of workstation with imaging work

distributed and shared amongst various workstations. The network usually includes a scanning station, a server and one or more editing, retrieval stations. A typical scanning workstation for a small, production-level project could consist of the following:

- Hardware (Scanners, computers, data storage and data output peripherals)
- Software (image capturing and image editing)
- Network (data transmission)
- Display and Printing technologies

2.11 Organization of Digital Images

A disc full of digital images without any organization, browse and search options may have no meaning except for one who created it. Scanned images need to be organized in order to be useful. Moreover, images need to be linked to the associated metadata to facilitate their browsing and searching. The following three steps describes the process of organizing the digital images:

2.11.1 Organize

The scanned image files into disc hierarchy that logically maps the physical organization of the document. For example, in a project on scanning of journals, create a folder for each journal, which, in turn, may have folder for each volume scanned. Each volume, in turn, may have a subfolder for each issue. The folder for each issue, in turn, may contain scanned articles that appeared in the issue along with a content page, composed in HTML providing links to articles in that issue.

2.11.2 Name

The scanned image files in a strictly controlled manner that reflects their logical relationship. For example, each article may be named after the surname of first author followed by a volume number and an issue number. For example, file name "smithrkv5n1.pdf" conveys that the article is by "R.K. Smith" that appeared in volume 5 and issue no.1. The file name for each article would, therefore, convey a logical and hierarchical organization of the journal.

2.11.3 Describe

The scanned images file internally using image header and externally using linked descriptive metadata files. The following three types of metadata are associated with the digital objects:

2.11.4 Descriptive Metadata: Include content or bibliographic description consisting of keywords and subject descriptors.

2.11.5 Administrative or Technical Metadata: Incorporates details on original source, date of creation, version of digital object, file format used, compression technology used, object relationship, etc. Administrative data may reside within or outside the digital object and is required for long-term collection management to ensure longevity of digital collection.

2.11.6 Structural Metadata: Elements within digital objects facilitates navigation, e.g. table of contents, index at issue level or volume level, page turning in an electronic book, etc.

2.11.7 Identification Metadata: Used for tracking different versions and editions of same digital work, i.e. pdf, HTML, PostScript, MS Word, etc. and TIFF, JPG, BMP, etc. In case of images. The simplest and least effective method for providing access is through a table of contents and links each item to its respective object / image. Content pages of issues of journals done in HTML would offer browsing facility.

Large scanning projects would, however, require a back-end database storing images or links to the images and metadata (descriptive / administrative). Back-end database used by most document management system holds the functionality required by most web applications. Important management systems like FileNet have now integrated their database with HTML conversion tools. These databases entertain queries from users through "HTML forms" and generate search results on the fly. Several digital library packages are now available as "open source" or "free-ware" that can be used not only for organizing the digital objects but also for their search and retrieval. Open source digital library software are elaborated in details in the section on "Components of Digital Library".

2.12 Conclusion

Digitisation is the process of converting the content of physical media (e.g., periodical articles, books, manuscripts, cards, photographs, vinyl disks, etc.) into digital format. In most library applications, digitisation normally results in documents that are made accessible from the web site, and thus are available on the Internet. Optical scanners and digital cameras are used to digitise images by translating them into bit maps. It is also possible to digitise sound, video, graphics, animations, etc.

Digitisation is the first step in the process of building digital libraries. Digitisation is also used for achieving preservation and archiving although

it is not considered as good option for preservation and archiving. It is highly labour-intensive and cost-intensive process that involves several complexities including copyright and IPR issues. However, digital objects offer numerous benefits in terms of accessibility and search. The documents to be digitised may include text, line art, photographs, colour images, etc. The selection of document need to be reviewed very carefully considering all the factors of utility, quality, security and cost. Rare and much in demand documents and images are selected as first priority without considering the quality.

The process of digitisation involves four steps, namely scanning, indexing, storage and retrieval. A scanned document is nothing more than a picture of a printed page. It can not be edited or manipulated or managed based on their contents. In other words, scanned documents have to be referred by their labels rather than characters in the documents. OCR (Optical Character Recognition) programs are software tools used to transform scanned textual page images into word processing file. OCR or text recognition is the process of electronically identifying the text in a bit-mapped page image or set of images and generates a file containing text in ASCII code or in a specified word processing format leaving the image intact in the process.

The quality of digital image can be monitored at the time of capture by four factors, namely i) bit depth / dynamic range; ii) resolution; iii) threshold; and iv) image enhancement. This section describes these parameters in detail. Image files are evidently larger than textual files. It is thus necessary to compress image files. Image compression is the process of reducing size of an image by abbreviating the repetitive information such as one or more rows of white bits to a single code. The compression algorithms may be grouped as lossless compression and lossy compression. The section describes compression technology and protocols.

An image scanning system may consist of a stand-alone workstation where most or all the work is done on the same workstation or as a part of a network of workstation with imaging work being distributed and shared amongst various workstations. The network usually includes a scanning station, a server and one or more editing and retrieval stations. A typical scanning workstation for a small production-level project, could consist of: Hardware (Scanners, computers, data storage and data output peripherals); Software (image capturing and image editing); Network (data transmission) and Display and Printing technologies.

Self-assessment Questions

1. Define digitization and describe the key components and features of digitization.

2. What is the criteria of selecting material for digitization? Also explain the various steps of digitization process.
3. Describe various file formats and media types of digitization, also explain the tools of digitization.
4. Why digitization is necessary? Discuss various techniques of digitization.

Activity:

1. With the help of your tutor develop a digitization plan of a public library having rare books and manuscripts collection.

Recommended readings:

1. Arms, W. Y. (1995 July). Key concepts in the architecture of the digital library. *D-lib Magazine*.

2. Arms, W. Y. (2000). *Digital libraries*. Cambridge, MA: The MIT Press.

3. Haigh, S. (1996). Optical Character Recognition (OCR) as a Digitization Technology. *Network Notes*, 37.

4. IMLS. *A Framework of Guidance for Building Good Digital Collections*. (n.d.). Retrieved from http://www.imls.gov/pubs/forumframework.htm

5. Jantz, R. (2001). Technological discontinuities in the library: Digital projects that illustrate new opportunities for the librarian and the library. *IFLA Journal*, *27*, 74–77. doi:10.1177/034003520102700207

6. Kenney, A. R., & Chapman, S. (1996). *Digital imaging for libraries and archives*. Ithaca, NY: Dept. of Preservation and Conservation, Cornell University Library.

7. Tyson, J. (2003 December). *How Scanners Work.* How Stuff Works. Retrieved from http://www.howstuffworks.com

8. Roelofs, G. (2005). *Multiple-image Network Graphics*. Retrieved March 2005, from http://www.libpng.org/pub/mng

9. Rothenberg, J. (1999). *Ensuring the Longevity of Digital Information*. Council on Library.

10. Sharma, A. (2001). *Digital Noise, Film Grain. Digital Photo Techniques*. Retrieved March.

11. Standard, D. (2003). *1 – Cataloguing and Metadata for Digital Images*. State Library of.

12. Standard, D. (2003). *2 – Digital capture, format & preservation*. State
13. Library of.

Unit. 3

BUILDING DIGITAL LIBRARIES: ROLE OF SOCIAL (OPEN SOURCE) SOFTWARE

Complied by: Muhammad Jawwad

Reviewed by: Dr. Pervaiz Ahmed

CONTENTS

INTRODUCTION

This unit is written to teach students after the emerging tools available on the Interment, blogging and social media. Blogging as a tool to outreach users in academic libraries is the main subject of this unit. It is an important component of digital libraries, that allows for timely two-way communication of information, bibliographic instruction, and news. It discusses the history and value of blogging. The unit will also focus on the Web 2.0 and Social Web approaches to digital libraries, analyses and discuss the integration of social annotations – uncontrolled user-generated content – into digital collection items. It briefly summarizes the value of annotations and finds that there is conceptual room to include user-generated content in digital libraries that they have been imagined as forums for social interaction.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is the significance of social media and social networking technologies in digital libraries?
2. What are emerging social media technologies and their applications in digital libraries?
3. How social media platform optimization strategies works in relation to digital library and information systems?
4. What legal requirements governing use of social media technologies in digital libraries?
5. What are the copyright and intellectual property protection practices and strategies in digital libraries?

3.1 Introduction

Evolution and revolution of digital information systems has influenced modern information organizations and professionals. New Internet platforms such as social media and social networking sites have become fundamental tools that enhance interpersonal communication. Social media and social networking technologies are powerful business tools used to provide, manage and support information and knowledge in digital libraries. An analysis of the modern technological trends in digital libraries reveal a greater adoption and use of different types of social media and social networking systems that also need copyright and intellectual property protection. Digital libraries provide different types of social media and social networking technologies that manage data, information, and knowledge in organizations. Information and knowledge management in digital libraries depend upon social media synergies that connect staff, clients, and supplies. The predominant goal of digital library systems is to develop and facilitate access to new information and knowledge. Information professionals and practitioners need to understand and appreciate the fundamental benefits of social media and social networking technologies in digital libraries.

Social media and social networking technologies are widely applied in the political, social and economic aspects of human life. Business, industrial and information organizations in equal measure utilize social media platforms for regular operations and the delivery of services. Fundamental aspects of analysis and discussion include modern social media and social networking technologies plus management of copyright and intellectual property protection in digital libraries.

3.2 Historical Background

Historically, the development of digital libraries began in the first decade of the 21st Century with many varied names and concepts used to describe the entities such as virtual library, electronic library and library without walls. The digital library is the collection, organization, preservation and management of information services and objects. Similarly, a digital library may be considered as a potentially virtual organization that comprehensively collects, manages and preserves for the long length of time rich digital content, and offers it to its targeted user communities. Digital library encompasses professionals from different fields such as data science, information sciences, information systems, human computer interaction and digital curation. Digital libraries are organizations that provide and manage electronic information and knowledge. Simply means modern digital libraries or computerization of traditional libraries.

Digital libraries provide endless access and unlimited opportunities to information and knowledge. In the digital library environment, social computing and

technological systems are used to manage and support information resources such as electronic books and journals (e-books and e-journals), bibliographic databases, websites, reference and citation management tools as well as digital theses and dissertations. The digital library is a geopolitical organization (borderless and boundary-less) with highly advanced technologies, hyper-connectivity and Internet-based platforms that provide diverse information services and resources. This is the library that applies technological systems and solutions to provide electronic information products and services beyond the physical borders and boundaries of the institution and organization. The digital library (also known as electronic or virtual library) is the multifunctional enterprise information system that provides and manages information products and services.

Social media is a means of communication through the internet that enables social interaction. It is an effective approach for people to use in communicating and interacting with each other. Simultaneously more organisations are applying social media to promote their services and interact with their users or patrons.

There are different variations of social media. The most popularly applied social media in digital libraries include blogs, Facebook, photo sharing, podcasts, RSS feeds, Twitter, YouTube, and wikis. There is some cross-classification as a few social media applications have more than one function. For example Twitter is a social networking site, but at the same time, it also serves to share photos and is classified as a microblog.

The marketing concept of "meeting the user where they are" becomes ever more necessary as users expect digital content to be readily available. As of 2021, 72% of American adults use at least one social media platform as described by Pew Research Center in 2021, and many digital libraries are turning to these sites to engage users, promote interaction with materials, and create relationships with similar institutions. Increasingly, these institutions are collaborating on social media campaigns to market their digital collections and expand their audience. While several of these efforts have gained significant attention, there is a lack of understanding as to how successful they are in meeting the goals of the participating institutions and how to quantify the benefits. Many social media strategies in digital libraries include goals for engaging users and making connections with other organizations.

Digital libraries exist as a space for users to access their holdings virtually, whether that be artwork, historical items, writings, or audio/visual materials. A hurdle many digital librarians face is reaching their intended userbase and engaging them with the materials, especially if there is a limited budget for marketing. Many businesses heavily incorporate social media into marketing strategies and use Twitter in particular to monitor and respond to user feedback. Digital libraries are also

adopting this strategy, using social media to track opinions, engage in conversation, and ask questions. However, failing to utilize these networking platforms to introduce library holdings to users is a missed opportunity.

Many user communities of book lovers and people interested in art, history, and other areas serviced by digital libraries have appeared organically on social media. Hashtags such as #shelfie, #bibliophile, #bookstagram, #museumlife, #artlife, and #library have hundreds of thousands of posts. The Metropolitan Museum of Art's Irene Lewisohn Costume Reference Library capitalized on the use of already-trending hashtags to draw attention to its niche collection, gaining 40,000 followers on Instagram in three years.

Library institutions can benefit from collaborating and creating a community on social networking sites. Schrier explains that, by making connections with similar organizations and visibly embedding itself in the community, a library gains credibility with users and comes to be known as a trustworthy and knowledgeable member of that community. Community, in this sense, involves activities such as sharing and commenting on other libraries' posts, helping to promote their events and new collections, and other simple social media interactions.

3.3 History of Blogging in Academic Libraries

Any history of blogging in academic libraries must take into account the development of two related concepts: Web 2.0, or Social Software, and Library 2.0.

3.3.1 Web 2.0 or Social Software

The concept of Web 2.0 originated as a list of characteristics of successful web properties. Among these are the Read/Write web, the web as platform, the Long Tail1, harnessing of collective intelligence, network effects, core datasets from user contributions, and lightweight programming models. Web users of the web engage many of these properties on a daily basis. The benefits of the principles of Web 2.0 and its technology are that they offer libraries many opportunities to serve their patrons better and reach out beyond the walls and websites of the institution to reach potential beneficiaries where they are, and in association with the task they happen to be undertaking. It is worth appreciating the level of integration and interoperability of Web 2.0 and Library 2.0 that are designed into the interface of a library portal or intranet.

Taking a cue from the Libraries and Social Software in Education (LASSIE) Project Report, we will use the term 'social software' for Web 2.0 tools and technologies, as these terms are broadly synonymous. As is evident from
its overall characteristics, social software is more about user-created content than

content created by an organization. Social software also includes development of user profiles and the use of 'folksonomies,' or tagging, to attach keywords users create to items to help them retrieve information.

Examples of key technologies that underpin the concept of social software are RSS feeds; blogs; wikis; social bookmarking and resource sharing, sites such as CiteULike, Del.icio.us; social networking sites, including MySpace, Facebook, and LinkedIn; media sharing sites such as YouTube, PhotoBucket, Flickr; and virtual worlds. The majority of these technologies are non-proprietary and accessible to all.

3.3.2 Library 2.0

The emergence of the Library 2.0 paradigm follows upon the emergence of Web 2.0. Because of the widespread use of Web 2.0 services, there are cultural changes affecting library users' information- seeking behaviours, communication styles, and expectations. As a result, the term 'Library 2.0' has been introduced into the professional language of librarianship as a way to discuss these changes, though what Library 2.0 is and what it means are still under discussion in the world of librarian blogging, the so-called 'biblioblogosphere.'

Library 2.0 is a concept wherein users are not only information consumers, but also content creators. It is a library without any boundaries and with the fullest participation of users as architects. In terms of the history of the use of the term 'Library 2.0,' Michael Casey first introduced it in September 2005.

The British Integrated Library Software (ILS) vendor, Talis, took an early interest in promoting the term. Michael Stephens exposed a wider audience to the term when he discussed Library 2.0 on the American Library Association's (ALA) Techsource Blog

Some of the popular definitions of Library 2.0 include the notions that

- Library 2.0 is the application of interactive, collaborative, and multimedia web-based technologies to web-based library services and collections (Maness, 2006).
- The heart of Library 2.0 is user-cantered change. It is a model for library service that encourages constant and purposeful change, inviting user participation in the creation of both the physical and the virtual services they want, supported by consistently evaluating services. What makes a service Library 2.0? Any service, physical or virtual, that successfully reaches users, is evaluated frequently, and makes use of customer input, is a Library 2.0 service. Even older, traditional services can be Library 2.0 if certain criteria are met. Similarly, the mere fact of being new is not

enough to make a service Library 2.0 if it fails to meet these criteria (Casey & Savastinuk, 2006).

- Library 2.0 is a concept of a very different library service, which operates according to the expectations of today's library users. In this vision, the library makes information available wherever and whenever the user requires it (Chad & Miller, 2005).

Given the above, we can describe Library 2.0 as a subset of library services designed to meet user needs caused by the direct and peripheral effects of Web 2.0 services, and leveraging concepts of the Read/Write Web, the Web as Platform, The Long Tail, harnessing of collective intelligence, network effects, core datasets from user contributions, and lightweight programming models.

This definition understands that Web 2.0 precipitates changes in user needs and that Library 2.0 services arise to meet these needs. The definition includes all implementations of Web 2.0 methodologies and technologies by libraries; however, this concept is not about replacing traditional technologies already in use, but rather about adding functionality.

Library 2.0 is a service model. The heart of Library 2.0 is user-centred change. It is a model for library service that encourages constant and purposeful change, inviting user participation in the creation of both physical and virtual services. It also attempts to reach new users and better serve current ones through improved customer-driven offerings. Each component by itself is a step towards better serving users. However, it is through the combined implementation of all of these components that Library 2.0 can be fully realized. The information environment within which libraries find themselves is changing, probably faster than ever before, which offers great opportunities for progressive libraries to reach out far beyond the boundaries of their buildings and websites, and to engage with an increasingly literate body of information consumers. The library domain has repeatedly evolved to embrace new technologies and to adapt in line with changing expectations, and it will doubtless continue to do so.

3.3.3 Blogging and Academic Libraries

Habib's model features blogs as one of the major social software platforms of Academic Library 2.0, which relies heavily on digital technology. Blogs are probably the most popular and best-known examples of social software. Jorn Berger coined the term "weblog" on 17 December 1997, but they have come to be known, most commonly, "blogs." People maintained blogs long before the term was coined, but blogging gathered momentum with the introduction of automated publishing system.

There has been a good deal of interesting literature generated on blogs and libraries in general and their application for academic libraries in particular. Laurel A. Clyde's book, *Weblogs and Libraries*, provides an extensive overview of the concept and process of blogging, lists a plethora of resources, and discusses various types of blogging software, strategies and tools for finding weblogs, and a detailed planning process for creating and maintaining weblogs, among other things.

3.3.4 What Makes Academic Library Blogging Valuable?

One of the most valuable aspects of a blog is that it can act as a Content Management System (CMS) to provide a platform for disclosure of information relevant in e-learning contexts. Academic librarians can assume the role of managing that content and be active in discovering and disclosing information relevant to academics, students, and the university community in general. Coulter and Draper, 2006, note another valuable aspect, arguing for the potential value of weblogs as community-building and learning tools in higher education, and reporting their use of blogs as a supplement to face-to-face information literacy instruction and as the sole means of library outreach to graduate distance learners. Zhuo, 2006, discussed the application of blogs in education in general and in academic libraries in particular, highlighting the contribution of blogs in imparting library instruction.

3.3.5 Blogs as Tool for Communication and Training

Caldwell, 2005 notes that blogging is an excellent way for library staff to communicate with each other and spread their message to a wider audience. Her interviews with some pioneers of academic library blogging suggested many positive outcomes, such as community building and breaking down the 'us versus them' barrier between librarians and patrons. The bloggers noted that the informal nature of blogging can contribute to the library community culture and that a library website is the perfect place to host blogs, especially for universities. Most importantly, blogging is a tangible way of demonstrating to the campus at large the commitment the library has to intellectual freedom and freedom of speech.

Libraries are reaping the rewards, too. Well-written, frequently updated public blogs help librarians relate to their patrons, generate support for new building initiatives, and market programs, collections, and services. As we noted above, internal staff blogs can also foster improved staff communication and camaraderie. It is evident from the extensive literature published in the context of academic library blogging that this initiative contributes to promoting and marketing library

services and resources in the digital age. Fichter, 2003, for instance, discussed the use of blogs for promoting and marketing library services. While librarians have seized the opportunities to market libraries in the real world via traditional media like newsletters, brochures, pathfinders etc., they can use blogs for similar purpose. By following simple principles like updating regularly, posting relevant information, and keeping posts short and crisp, librarians can maintain their blogs and promote their library events, resources, and services while ensuring that readers return to their blogs on a regularly basis.

Another use for blogs, in terms of marketing and service promotion, is to raise awareness of library services to on- and off-campus users. Many of the academic library blogs maintained at universities such as the University of Minnesota's UThink Blogs, Fairfield University's Your Voice Counts! (20082008Your Voice Counts! 2008!!), Georgia State University's University Library Blogs deliver current awareness service. Another area in which blogs can serve as potential tools is in delivering digital current awareness service. A blog is a tool through which communities of information-seekers and information providers can collaborate. Blogs act as organizers of data; each element in a blog is a standard data object that can be referenced. Completeness and accuracy are the traditional measures of success of the reference transaction.

Blogs can also be a tool for providing information literacy instruction. For example, Stuart Boon and Sheila Webber, both of the Department of Information Studies at the University of Sheffield, set up the Information Literacy blog. What the blog offers is evident from its title, and the posts are targeted to a wide audience, including librarians, faculty, and non academics. Boon believes that blogs sit at the heart of the librarian's role.

3.6 Blogging as Part of Alternative Models of E-Learning

From the literature, we see that blogs can be used for promoting and supplementing e-learning. Although social software tools like blogs are not created primarily for

educational purposes, they can be easily adapted to support learning. Using social software can help facilitate an approach to e-learning, which differs from using a Learning Management System (LMS), as Dalsgaard, 2006, notes. The personal tools and social networks support a self-governed, problem-based, and collaborative learning process.

According to Glogoff, 2005, instructional blogging operates as a knowledge-centred instructional tool. In this model, the instructor involves students in research activities, engages them in discussion with practitioners, and leads them through developmental concepts of the discipline's knowledge domain. As a valuable e-learning tool, blogging can be used in a number of ways to engage students in discussion, exploration, and discovery.

It is appropriate for both hybrid and fully online courses. Glogoff, 2004 explains that blogging is a simple and adaptable technology that can be used to construct learning environments that work well with receptive, directive, and guided discovery instructional techniques. Through their empirical study, Williams & Jacobs, 2004, concluded that blogs have the potential, at least, to be a truly transformational technology in that they provide students with a high level of autonomy while simultaneously providing opportunity for greater interaction with peers. They suggest that blogging tools would be a valuable addition to any LMS.

3.3.7 Blogging as Part of Effective Knowledge Management and Sharing

Weblogs serve as personal publishing tools for knowledge workers in the digital world and support the process of knowledge work. Besides being personally beneficial to the knowledge worker, blogging also helps in disseminating knowledge throughout an organization.

Basing his assessment of knowledge work processes on the Framework of Knowledge Work analysis developed by Efimova, 2004, Röll quoted a relevant passage from her article, which is very much relevant to the functions mostly performed by the academic librarians in the digital world: "Much of the work of

finding, interpreting and connecting relevant pieces of information, negotiating meanings and eliciting knowledge in conversations with others, creating new ideas and using them to come up with a final product, happens in the head of a knowledge worker or as a part of communication or doing work."

Röll concluded that blogs support knowledge work by providing a space to capture information, annotate it, reflect on it, get feedback, share, discuss and network with others. They also provide organizational benefits. Blogging can facilitate corporate/organizational story telling, an ace leadership and motivational technique and a powerful concept for knowledge sharing in organizations. Since blogs have a human voice, they encourage interaction with the author and other readers, as well as providing context a point well noted by Nichani & Rajamanickam, 2001.

In light of all this research, we maintain that the research and literature provide a substantial base for the assumption that academic librarians are equivalent to knowledge workers, and that blogging helps organize their work to better manage and share knowledge.

3.3.8 Blogging as a Means of Reaching Out to Distance Learners

Today's increasingly e-learning culture collapses the boundaries between regular on-site students and off-campus, external programme students, i.e., distance learners. Hence, academic librarians are playing the roles of instructors and knowledge workers by reaching out to them with the help of social networking tools like blogs.

There is a body of literature and research that demonstrates that blogging is an excellent means of reaching out to distance learners. Secker & Price, 2007, for example, described a recent project funded by the University of London to explored how social software of Web 2.0 technologies can enhance the use of libraries by distance learners.

LASSIE, funded by the Centre for Distance Education, University of London, was a 9-month project completed in December 2007. The project addressed concerns at the University of London that the external programme students, i.e., their distance learners around the world, were not making full use of electronic library resources.

The reasons for this were thought to be varied, including low awareness of how to access library resources and information literacy issues. The project explored how the provision of appropriate social software might impact the use of library services by distance learners.

Their exploration was a four-step project plan:

1. Conduct a literature review on physical libraries as social spaces;
2. Explore the current use of social software in libraries;
3. Evaluate the potential of social software in libraries;
4. Identify specific external programme courses to pilot social software tools and evaluate their impact on student learning, using an email questionnaire and interviews to gather qualitative data.

The LASSIE Project performed an extensive literature review that focused on topics such as libraries and distance learners, libraries as social spaces, Web 2.0, Library 2.0, overview of Web 2.0 technologies, and libraries using social software.

Developing social software for blogging will highly benefit off-campus users by helping them access library resources remotely. Much as learning environment are changing in on-campus programmes, where students and faculty are increasingly employing virtual and e-library services on their desktops and laptops, the same technology can be used and adapted for distance learners. Indeed, we suggest that distance learning institutions should begin to think along these lines and consider providing effective library support to distance learners. The possibility of maximizing the use of social software in a developing country such as India provides some fascinating alternatives.

Though guidelines for providing library services to distance learners exist, whether from Association of College and Research Libraries (ACRL), American Library Association (ALA), Commonwealth of Learning (COL) topical guide and the Indian Library Association (ILA), we argue that precise policies need to be laid down for initiating and implementing them.

3.4 Conclusion

We have looked at a number of items in this unit. First, we discussed the concepts of Web 2.0 and Library 2.0. Then we examined social software and its relations to Library 2.0, especially Academic Library 2.0. Next, we looked at a high-level overview of blogging and academic libraries, and the value of blogs in a number of contexts. After that, we outlined several examples of how blogs are used in academic libraries, whether as new tools to offer traditional services, as part of alternative models of e-learning, an element of effective knowledge management and sharing, and as a means of reaching out to distance learners. Our hope, in our review of the research and literature, is that we have shown that blogging is a valuable tool for academic librarians.

Self-assessment Questions

1. Discuss the significance of social media and social networking technologies in digital libraries.
2. What emerging social media technologies are applicable in digital library? Discuss in detail.
3. Describe the historical background of development of digital libraries and impact of social media strategies.
4. Discuss the role of web 2.0 and library 2.0 in the marketing of digital library contents and development.
5. Explain the role of blogging in academic libraries.

Activity:

1. Prepare an alternative model of e-learning through blogging as part of effective knowledge management and sharing.

Recommended Reading:

1. Bar-Ilan, J. (2007). The use of Weblogs (blogs) by librarians and libraries to disseminate information. *Information Research, 12*, Paper 323. Retrieved October 17, 2008, from http://InformationR.net/ir/12-4/paper323.html

2. Bell, S. (2005). Where the Readers Are. *Library Journal*, 8-12. Retrieved February 17, 2009, from http://nrs.harvard.edu/urn-3:hul.eresource:abiinfor

3. Boon, S., & Webber, S. (2009). *Information Literacy Blog*. Retrieved February 18, 2008, from http://information-literacy.blogspot.com/

4. Boutin, P. (2008). Twitter, Flickr, Facebook Make Blogs Look So 2004. *Wired Magazine*. Retrieved December 3, 2008, from http://www.wired.com/ entertainment/theweb/magazine/16-11/st_essay

5. Brookover, S. (2007). Why We Blog. *Library Journal, 132*, 28-31. Retrieved 2/18/2009, from http://nrs.harvard.edu/urn-3:HUL.ejournals:sfx954921392997

6. Bryant, T. (2006). Social Software and academia. *EDUCAUSE Quarterly*, *2*, 61–64.

7. Caldwell, T. (2005). Pen a Blog Buster. *Information World Review*, 16-17. Retrieved 2/17/2009, from http://nrs.harvard.edu/urn-3:hul.eresource:abiinfor

8. Efimova, L. (2004). *Discovering the iceberg of knowledge work: A weblog case.* Paper presented at the OKLC 2004. Retrieved January 25, 2008, from https://doc.telin.nl/dscgi/ds.py/Get/File-34786

9. Fialkoff, F. (2005). The Power of Blogs. *Library Journal*, *130*(6),

10. Ficther, D. (2003). Why and How to Use Blogs to Promote Your Library's Services. *Information Today, 17*. Retrieved October 17, 2008, from http:// www.infotoday.com/mls/nov03/fichter.shtml

11. Glogoff, S. (2005). Instructional Blogging: Promoting Interactivity, Student-Centred Learning, and Peer Input. *Innovate: Journal of Online Education, 1*. Retrieved February 19, 2008, from http://innovateonline.info/Gordon, R. S. (2005). Revenge of the NextGen People. *Library Journal*, *130*(9), 78–78.

12. Gorman, M. (2005). Revenge of the Blog People! *Library Journal*, *130*(3), 44–44.

13. Reichardt, R., & Harder, G. (2005). Weblogs: Their use and application in science and technology libraries. *Science & Technology Libraries, 25*, 105-116. Retrieved February 18, 2009, from http://nrs.harvard.edu/urn-3:HUL.ejournals:sfx958480310487

14. Röll, M. (2004). *Distributed KM – Improving knowledge workers' productivity and organizational knowledge sharing with weblog-based personal publishing.* Paper presented at the Blog Talk 2.0, "The European Conference on Weblogs.

15. Roper, T. (2009). *Tom Roper's Weblog*. Retrieved February 18, 2009, from http://tomroper.typepad.com/

16. Sauers, M. P. (2006). *Blogging and RSS: A Librarian's Guide*. Newark, NJ: Information Today.

17. Schrecker, D. L. (2008). Using Blogs in Academic Libraries: Versatile Information Platforms. *New Library World*, *109*(3/4), 117–129. doi:10.1108/03074800810857586

18. Secker, J. (2008). *Case Study 4: blogging and libraries*. Retrieved February 17, 2009, from http://clt.lse.ac.uk/Projects/Case_study_four_report.pdf

19. Secker, J., & Price, G. (2007). Libraries, social software and distance learners: blog it, tag it, share it! *New Review of Information Networking*, *13*(1), 39–52. doi:10.1080/13614570701754536

Unit. 4

INFORMATION PRESERVATION AND SERVICES, DIGITAL PRESERVATION CHALLENGES

Complied by: Muhammad Jawwad

Reviewed by: Dr. Pervaiz Ahmed

CONTENTS

INTRODUCTION

This unit is developed to provide students an understanding about Information Preservation and Information Services in the Digital Age. It will discuss the strategies of storage and preservation. Information services in the digital age and their benefits have also been discussed. This unit will outline the challenges of digital preservation, infrastructure needed in preservation and reference model for preservation. Also discussed are the threats and challenges to preservation. There is a description of the implementations of tools and infrastructure components to support repositories in their task of long-term preservation of digital resources, including the capture and preservation of digital rights management and evidence of authenticity associated with digital objects. Digital Reference and Information Services in Digital Environment," will also be explained with an important concept, information literacy which assumes importance in the era of digital information. Digital reference service, issues in digital reference service are also discussed. Moreover, web enabled reference services and marketing digital services also highlighted in the unit.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is information preservation and how?
2. What are the purposes, advantages, and disadvantages of digitization?
3. How technology help in preservation of various library material?
4. What are strategic benefits and strategies of digital preservation?
5. What type of reference and information services can be provided in digital environment?
6. What challenges and threats are faced by libraries in digital preservation?

4.1 Introduction

As developed for more than 200 years, academic libraries have generally been designed first and foremost as a place to collect, access and preserve
print collections. To enter and use them was considered privilege. Despite their handsome exteriors the interior spaces were often dim and confining.
The buildings were difficult to navigate and specialized services and collections were inaccessible to all but the serious scholar. Building planning and design of these libraries were primarily devoted to the preservation and security of materials and to the efficiency of the library collection services.
Prime space was routinely reserved for processing materials.

Information today is being produced in greater quantities and with great frequency than at any time in history. The ease with which electronic
information can be created and published makes much of what is available today. Digital is now often the first choice for creating, distributing and
storing contents from text to motion pictures to sound. As a result digital content embodies more and more of the world's intellectual, social and
cultural history. And the preservation of such content has become a major challenge for society.

Libraries collect and preserve books and other materials for future generations to ensure that every citizen has equal access to information.
With the advent of the internet and World Wide Web libraries can extend their reach unbound by time and space. The internet had made shared
knowledge and technical collaborations across national boundaries a viable endeavour. This is a defining moment for libraries. Technological innovation and the ubiquity of communication tools, economic uncertainty, changes in workplace and educational structures, the globalised economy, generational differences, the blurred distinction between the production and consumption of information and heightened national security are just some of the factors affecting the creation of digital library programs. In addition, there is an almost insatiable demand for content to meet the needs of the more than
six billion internet users worldwide. And libraries no longer have market concerned on information services. Studies have shown that todays' students
first turn to internet and that many library patrons are willing to settle for less, convenience over comprehensiveness.

Academicians, students or researchers, all have a craving for information. Digital library is being renamed to library itself because everything in this library is stored electronically and digitally. Digital libraries are sets of electronic resources and

associated technical capabilities for creating, searching and using information. In a sense they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium [text, images, sounds, static or dynamic images] and exist in distributed networks. The content of digital libraries includes data, metadata, that describes various aspects of the data and metadata that consists of links or relationship to other metadata, whether internal or external to the digital library

4.2 Digital Libraries

Digital libraries are constructed- collected and organized by a community of users and their functional capabilities support the information needs and the uses of that community. They are a component of communities in which individuals and groups interact with each other using data, information
and knowledge resources and systems.

In this sense they are the extension, enhancement and integration of variety of information institution as a physical place where resources are selected and collected and organized, preserved and accessed in support of user community. Implicit in its definition digital library is a broad
conceptualization of library collection. The meaning of digital library is less transparent than one might expect. The words conjure up images of cutting-edge computer and information science research. They are invoked to describe what some assert to be radically new kinds of practices for the management and use of information.

According to Digital Library Federation, “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 collections of digital works so that they are readily and economically available for use by a defined
community or set of communities.”

4.3 Purpose of Digital Libraries

Amply has been argued about why to digitize and what to digitize. The main cause for it is optimum access and accurate use. Though there is wide and huge scope of debate on the purpose of digitization many libraries across the globe have adopted this culture. These libraries have seen something very beneficial and lucrative about such project and decided to go with it. And most of the goals achieved through

digitization have ensured availability, identity, understandability, Fixity, authenticity, viability and renderability of digital information.

Availability- A library that wants to preserve digital images it created in a scanning project is likely to have preservation masters in its possession, perhaps offloaded to tape or on DVDs. A library that wants to preserve the intellectual output of a university will have much hardertime and may need to work with the faculty and administration to establish institutional repository. Deposit agreements, license negotiated to provide a library with an archival copy and contracts with publishers are all ways to get copies of published materials. Depending on the materials and circumstances getting a copy of the objects may be long or quite difficult.

Identity- if the end of digital preservation is long term access and usability, the digital objects must be described in sufficient detail and should carry descriptive metadata within them. Many contemporary file formats support embedded metadata, but only if object creators take advantage of their capabilities. Whether the metadata is internal or external there are no separate standard for descriptive metadata for preservation.

Understandability - A repository must ensure that the preserved information in independently understandable to its user community. For example descriptive metadata may tell us that a dataset represents the results of a certain pre lection poll, but unless we have the codebook we won't know what questions were ask and how the answers are represented in the file. The repository is responsible for providing and preserving enough information as metadata documentation and related objects to enable future users to understand the presented objects.

Fixity- Preservation systems must protect digital objects form unauthorized changes, whether deliberate or inadvertent. Industry standard computer security regimes are the best defence against both malicious and careless behaviour. This includes virus protection, firewall, tight authentication, intrusion detection and immediate attention to security alerts. Media degradation may also cause bitstream corruption and is prevented by sound storage management practices, including climate control and media refreshment.

Authenticity- A preservation program may not be able to guarantee that all the digital objects it handled are authentic, but preservation treatment should not compromise the authenticity of the object in any way. There must be policies and procedures to endure data integrity and to ensure that the chain of custody and all authorized changes are documented. The event history pertaining to a digital object is known as its "digital provenances" and is a critical part of preservation metadata.

Viability- Viability is a quality of being readable from media. Media deterioration and media obsolescence are threats to viability, and both of these are experienced by most of us in our everyday lives. Viability of a digital file is easily ensured when the files are actively managed, as digital data, then unlike analogue data can be copied without loss. Files should be copied periodically to new media and back up copies should be stored on different physical devices.

4.4 Digital Storage and Preservation Strategies

Digital files can provide extra ordinary access to information. They can make the remote accessible and hard to see visible. Digital surrogates can bring together reserve materials that are widely scattered about the globe allowing viewers to conflate collections and compare items that can be examined side by side solely by virtue of digital representation. Digital technology can also make available powerful teaching materials for students who would not otherwise have access to them. Among the most valuable types of materials to digitize from classroom perspective are these from special collections of research institutes including rare books, manuscripts, musical scores and performances, photographs and graphic materials and moving images. Often these items are extremely rare, fragile, or in fact unique and gaining access to them is difficult.

Digitizing these types of primary sources offers teachers at all levels unheard of opportunities to expose to their students.
Image processing – the manipulation of images after initial digital capture can greatly expand the capacity of the researcher to compare and contrast details that the human eye cannot see unaided. Images can be enhanced in size sharpness of detail and color contrast. Through image processing a badly faded document can be read more easily, dirty images can be cleaned up and faint pencil marks can be made legible. While we know that the daily number of hits at the site of Wikipedia is greater than the number of readers who visit the library reading room each day, we have very little data now as to how much these types of online images are used and for what purpose. Some large libraries are attempting to compile and analyze usage statistics, but this intensive task presents quite a challenge. We need more user studies before we can assert confidently what may seem self evident to us now.

Adding digitized special collections to the mass of information available on the internet is in public interest and enhances education. We also need to ensure that libraries are working collaboratively in their efforts to digitize materials so that together they can create a mass of research sources that are complementary and not

duplicative and they begin to fulfill the promise of coordinated digital collection building.

However at present there is no central depository providing information of what has been digitized and with what processes, as there is for titles that have been microfilmed for preservation. Some of the drawbacks of digital technology for access, as for preservation stem from the technology's uncanny ability to represent the original in a seemingly authentic way. Working with digital surrogates can distort the research experience somewhat by taking research materials out of context of the reading room. The nature of computer display makes only serial viewing possible, very different indeed for example from spreading photographs in their original sizes around a flat surface and looking at them simultaneously and in different groupings. Every page, every object is mediated by the screen which automatically flattens the images and removes their contexts.

And a digital image no matter how high the resolution and sensitive the display monitor is always presented through the high density of information of the computer screen compromising the high density nature of analogue materials which can be artificial for assessing some visual evidence.

Digital raw materials on the web are not as raw as they might appear to be. Many of the items that may be viewed on the websites of renowned institutions come from special collections that are large, often catalogued only at the collection level and often unedited with few descriptions that aid a scholar. The amount of physical preparation and intellectual control work that is needed for every digital project is very large.

Scanning is very expensive process, and most of the cost occurs before the item is being laid on the scanner. Part of that is physical preparation of research into and description of an item. The collections that are on the web are in an real sense publications accompanied as they are by a great deal of descriptive information created in order to make the items understandable in the context of the internet.

4.4.1 Access

Despite high cost of digital conversion many institutions have taken ambitious projects in order to find out for themselves what the technology can do for them. The impact of digitizing projects on an institution, its way of operating its traditional audience and its core functions are often hard to anticipate. The challenge of selecting the parts of large collection that will be scanned is, for some novel task that calls into question principle of collection development and access

policies. Making information available on the internet removes the very barriers from use that we take for granted in physical collection.

There are ways to build in electronic barriers to access for all or portions of site, using much of the same technology that commercial entities use in granting fee-based access.

However, constructing these barriers add a layer of administrative ccomplexities to manage the site that libraries and archives may not be prepared to take on, even if the technology does exist. Only when digitization is viewed specifically as a form of publishing and not simply as another way to make resources available to researchers are the thornier issues of selection for conversion put into an editorial context that provides a strong intellectual and ethical basis for imaginative selection of complex materials.

4.4.2 Ownership

Many of the collections that may be of the highest research and teaching value will not be digitized for web access because of the structure of copyright that may apply. For this reason library websites these days contain a disproportionate amount of public domain material which distorts the nnature of the source base for research restricted to the web.

4.4.3 Strategies

Preservation strategies can be thought of as falling into two categories. The methods in the first category address the goals of fixity and viability and include techniques such as copying data to new media of the same type, copying data to the newer media, and maintaining multiple, frequently verified copies of data. These activities are often referred to as "bit level "or "passive" preservation.

The methods in the second category attempt to address the goals of renderability and authenticity and are unique to the preservation realm.
Format migration and emulation are often touted as the two main strategies for digital preservation. In fact numbers of different strategies are available to preservationists and multiple approaches are often used together to good effect. When strategies addressing renderability are employed it is called "full" or "active" preservation. So the main strategies for preservation include, Technology preservation, Emulation, Universal Virtual Machine, Universal virtual computer, Format Migration, Format Normalization.

- **Technology Preservation--** Often called the "computer museum approach" technology preservation is familiar to anyone who still owns a record player

for listening to vinyl LPs. If a format depends upon a particular combination of hardware and software for rendering, it should be possible to preserve at least a few working examples of the obsolete platform. Technology preservation is generally considered an interim approach at best because it is not scalable. However, preserving old technologies can provide historical information about genuine behaviours of obsolete application. There is a fairly strong consensus that technology preservation is probably impractical as a digital preservation strategy, largely because the number of hardware components requiring preservation would soon grow to unmanageable levels while the cost of maintaining preserved hardware in working order would become increasingly expensive as the hardware aged. Moreover, hardware as a physical object requires physical space for storage which would significantly increase the cost of preservation. On top of this there would be additional costs involved in preserving the software intended to run on the preserved hardware as well as digital objects the preserved hardware and software are intended to support. In short technology preservation requires the preservation of every component of the original computing environment along with the archived content itself. While in some cases it is desirable – for example a museum exhibit designed to offer visitors an authentic computing experience from earlier times- the costs are too high to warrant serious consideration of technology preservation as a strategy for maintaining access to digital objects over long.

- **Emulation--** It involves the use of hardware and/ or software that allow computer instructions written for one platform to be run on another platform. Emulation has been in use in the computer industry for years to extend the life of programs written for earlier models of machine. Today emulation is widespread, particularly to allow programs written for one microcomputer operating systems to run on any other. Emulation as a strategy for long term digital preservation however is still largely experimental. Many experts consider that of all digital preservation strategies, emulation harbors the highest potential costs. First a library of emulators must be developed and maintained- of course the greater the number of formats supported by the repository, the larger the library of emulators needed. Second the software environment for the archived digital objects i.e. operating systems and application program, need to be preserved as digital objects in their own right, with all of the associated costs of doing so. And finally the emulators themselves represent digital object subject to the vagaries of an ever-changing technological environment: as current technologies are eventually displaced and become obsolete. New emulators

must be written that are compatible with whatever new environment emerges.

- Universal Virtual Machine-- Universal virtual machine behaves as intermediate layer between the emulator and the current platform isolating the emulator from the technology changes. One of the problems with simple emulation is that modern computerer technology is moving target- not only does an emulator has to be written for each obsolete platform, but emulators must be updated or rewritten as a current platform change. Although the UVM itself may require updating or rewriting for new platform, it is presumably less work to update one UVM than dozen of emulators.

- Universal virtual computer--Raymond Lorrie of IBM expanded the concept of a Universal Virtual Machine to that of a Universal virtual computer for preservation. In this approach, files of a given format are translated to a simpler Logical Data View by a "decoder" program written to run on the Universal Virtual Computer. The original file, the Logical Data View and a schema describing the Logical Data View are all archived together. In the future, files in the format can be rendered by the first building a UVC emulator to run on then current hardware then executing the decoder to generate the Logical Data View and finally writing a viewer to render the Logical Data View according to the schema. National Library of the Netherlands has been a leader in exploring the use of the UVC in a production preservation environment.

- **Format Migration--** Format migration is also called "forward migration" that creates a version of a source file in a different format that is considered to be a successor format. This is routinely done by common desktop application such as Microsoft Word or excel, which can open a file written by an earlier version of the program and save it in the current format. In some cases the successor to one format may be an entirely different format. One concern about the use of format migration for digital preservation is the likely need for successive migration over time. Since any format transformation could potentially lose or even add information, it is possible that successive migrations would accumulate errors leading to results less and less like the original. A counter strategy is to save the original and write programs to migrate directly from the original to the current format. In particular format migration involves an alteration in the way the 'ones and zeroes' of a digital object are encoded usually in order to make the object's bit streams accessible and interpretable by contemporary software/

hardware environments. The cost of format migration hinges on a number of factors, including the frequency of occurrence, the availability of standardized tools for carrying out the migration, and the tolerance of information loss sustained as a by-product of the migration process. Clearly, the more frequent the need to migrate archived objects to new formats, the greater the cost. Unfortunately, this factor is largely outside the repository's control, and lies instead in the hands of the pace of innovation and user expectations. Standardized tools limit the costs of migration, and repositories can control this factor, to the extent that they can choose formats for which such tools exist: however, the flexibility of choices in this regard may again be limited by the expectations of users. Similar reasoning applies to the degree of tolerable information loss, which can be inversely related to the costs of migration, but also governed by the expectations of the user community.

- **Format Normalization--** The process of format normalization creates a version of a source object in a preferred format while maintaining the essential properties of the original. For example, textual documents in proprietary word processing formats could be converted to Rich Text Format or to an open XML based format. Some preservation systems, particularly those designed for archival materials, normalize all incoming documents on ingest. This has an advantage that there are fewer formats for the repository to support and maintain over time. The disadvantage is that normalization can be loss and unless the original is also preserved, the initial decision as to what properties must be maintained is critical. Normalization can impact structure, appearance and functionality in widely differing ways, depending on the original format of the object.

4.5 Advantages of Digitization

In recent years, a growing understanding of the costs of digitization, in terms of both time and financial resources, has placed a greater focus on developing digitization initiatives and programmes that will realize tangible and strategic benefits for the institution and its users, rather than opportunistic or short-term projects that are limited in their scope or focus. Consequently, it has been necessary to articulate clearly the concrete benefits of running digitization projects at the outset.

Digitization is a complex process, and there are concrete benefits to be realized from many types of digitization projects. These can be summarized as Access, Support of preservation activities, Collection development, Institutional and strategic benefits, and Research and education.

4.5.1 Access--The primary and usually the most obvious advantage of digitization is that it enables greater access to collection of all types. All manner of material can be digitized and delivered in electronic form and the focus of the content that is selected for digitization varies across institutions. Digital materials can be made available to a broader audience than those who have the resources or ability to travel to see the analogue collections and access can be expanded to non- traditional audiences such as lifelong learners. Audience can access the collection for often unanticipated and broad-ranging research interests. Whatever the audience, their access to the materials is enhanced by the advantage of the digital format. With the application of the right technological tools, and careful attention to the design of the user interface, it is possible to search, browse and compare materials in useful and creative ways. Patrons may scroll or browse through thumbnails of the materials in image catalogues, including images of materials that were previously inaccessible, such as glass plate negatives, or over sized or fragile materials. Digital images or texts can be integrated with, and linked to other materials, to provide an enriched archive of materials.

4.5.2 Support of Preservation Activities—Developing a digital surrogate of a rare or a fragile original object can provide access to users while preventing the original from damage by handling or display. This was the motivation behind the digitization of many priceless artifacts. Often, the fragile condition of collections prevents or reduces their use. Digitization is not a substitute for traditional preservation microfilming, however. The digital format is too unstable, and issues related to the long-term preservation of digital media have not yet resolved.

4.5.3 Collection Development--The provision of digital materials can overcome gaps in existing collections. Primarily, there is an opportunity for collaborative digitization initiative to allow the re-unification of disparate collections. It is often the case that materials that were originally part of a complete collection are now held in far-flung locations, and there is a growing desire to present at least a virtual sense of what the entire collection would look like. Many projects have been motivated by the goal of virtually 're-unifying' such materials. Digitization is a means of creating resources that can be re-purposed for unforeseen uses in the future. Changing research trends may alter the demand for items in a collection: the development of new fields of study means that collections once perceived as ephemeral or of low research value are now heavily researched. Similarly, collections of items that were once in high demand are now banished to offsite storage for lack of use. Furthermore, libraries are increasingly under pressure to provide access to materials in response to user requests and are transitioning policies from collecting material 'just in case' someone will need it, to one of

developing relationship which allow the library to deliver material from elsewhere 'just in time' to answer a user's need.

4.5.4 Institutional and Strategic Benefits—There is no doubt that digitization programs can raise the profile of an institution. Projects to digitize priceless national treasures or valuable scholarly materials, if done well, can bring prestige to the whole institution. Raising the profile of an organization by showcasing digital collections can be useful public relation exercise. Digital collection can also be used as leverage with benefactors and funders by demonstrating an institutional commitment to education, access, and scholarship. Certain funding opportunities exist for digitization and it may be expedient for an institution to use them as an opportunity to accelerate a digitization program.

Developing digital projects can have long-term benefits for the institution, although it may take many years to realize these benefits fully. Such initiatives may create an opportunity for investment in the technological infrastructure and can create an opportunity to develop the overall technological skill base among staff. Staff members may benefit from access to digitization programs that give them an opportunity to learn about new technologies, that way it gives tremendous opportunities for staff development. Many funding opportunities are contingent on collaborations and partnerships between several institutions, so this can be an excellent opportunity to develop strategic liaisons with other institutions.

4.5.5 Research and Education-- Digitization of cultural heritage materials can have tremendous benefits for education. Many institutions present educational 'modules' on their websites, presenting 'packages' of educational material based around their collections. Museums have been particularly successful in this respect, as most organizations have in-house educational departments, which have been charged with developing materials that will exploit the potential of technology for delivering educational resources to all levels of learners.

The advantages to academic research and advanced scholarship are equally impressive and the potential of networked technologies to create a dynamic reading and scholarly environment is driving digitization initiatives at many institutions. Digitization can also be the first step in conducting advanced research on historical materials. Ancient documents present a prime candidate for digitization because of their historical import, combined with centuries of exposure and degradation.

4.6 Reference and Information Services in Digital Environment

The motto of providing information services is not at all changed. But as the

environment is switched to the digitization, the service staff is required to be trained accordingly. The digital library offers users the prospect of remote, around-clock access to electronic resources. There is serious need for bringing service perspectives and traditions of the physical library to the digital library.

Human assistance in the digital environment is badly needed, as users lack the most basic skills using web-based catalogues, indexes, circulation, reserves and ILL systems search engines and databases. This new digital service environment requires that we radically change our perspectives on user needs and transform the ways in which we organize roles of reference staff to serve these needs. There are some workable progressive reference trends to implement in digital library service atmosphere as: tiered service, roving librarians, growth of research consultation activities, increasing number of professionals working outside of the library, use of more paraprofessionals etc.

The terms public services and technical services are ingrained in our collective culture. These two terms certainly embody a dichotomy. This mild opposition has been rendered even milder by recent coinage such as "access services", "information delivery services' and interpretative services. The two services have been psychologically divided for many years and this has been detrimental for services to users and for the quality of work life of librarians. " Public services " seems to imply groups of people who are uniquely suited to interaction with the users of the library, whereas "technical services" denotes a group of secretive, hidden librarians, devoted to the area of cataloguing and the dark world of systems.

However, library collections are defined, they are either inert or randomly used without the human interaction that we call reference service. Person to person reference service is on its way out that will go the way the Library of Congress catalogue card and readers' advisory services. As with many other predictions concerning "virtual libraries" and the like, forecasting the death of face to face reference seems to ignore the manifest advantages and popularity of service, it seems that one has to have extremely strong argument to facilitate or allow the demise of a service that is both expected and appreciated by a wide range of library users.

Knowing the reference collection well is important to good reference work, but so is an intimate knowledge of the wider collections. If we can use technology and electronic collections to enhance this complex structure, then so much the better for all. There are those who propose that technology can be employed to provide a satisfactory alternative to the nuances of the interaction between librarian and user, the librarian's familiarity with range of the recorded knowledge and information, and the subtleties of information and knowledge seeking.

Though reference librarian is not always directly involved in preservation of the records of civilization, he is vitally concerned about the totality of the record. As far as electronic resources are concerned, the reference librarian has a duty to view them in the light of all other resources. This means using them when they are the best source and eschewing them when they are not.

Good reference librarians are aware of and value the whole world of recorded knowledge and information- from books to maps, videos, electronic resources and everything in between. They are concerned for all resources of all kinds and their onward transmission to posterity. They cannot therefore be indifferent to the fact that the inchoate nature of electronic resources and their mutability poses a preservation problem unlike anything in the last 500n years. There is a very real chance that much of what is available electronically will not be available in few years. By unavailable means lost forever, not just difficult to find. This is a sea change in the history of communication or rather a reversion.

4.7 Skill Requirements

The proliferation of knowledge of IT in general and computer science in particular has reached a mad level over last two decades that most of LIS courses incorporated too much of it in their curricula without matching provision for wards to acquire skills. The academic and theoretical aspects like history of computers, generations of computers, etc. were taught on par with any computer science course and the utility of which in practice was found negligible.

Interestingly some concepts and techniques of librarianship have again surfaced with new names in the latest IT and electronic world. For example, push or feed technology is conceptually same as SDI and Meta data is same as bibliographic data or surrogates of information sources. One undesirable feature in the training of LIS professionals in the areas of classification and cataloguing till 70s and in the areas of IT or computer science during last two decades is that there has been lopsided emphasis on specific schemes, systems, soft wares or languages, and general underlying principles, concepts and techniques are not given due importance.

Unfortunately, that even principles and techniques are enunciated as corollaries to a given scheme or system. The electronic environment of 21st century will encompass a wide range of technologies including computer, communication, storage, recognition and other technologies. As such it is easy to say that knowledge and operating skills in all these areas are required by future LIS professionals. As mentioned earlier, as for as breadth and scope of required skills are concerned, LIS

professionals must have technical skills, IT skills and managerial skills. Before getting into these three broad groups of skills, we may note that skills are not generally acquired by self study or listening to lecturer. What are called 'practical sessions' in the traditional schools of library and information science also hardly impart skills. At the best, they arouse curiosity for knowledge.

Secondly, when we discuss skill requirements that too in alien areas like management and IT, it is necessary to be clear about the level or depth of the skills expected of LIS professionals. As far as IT related skills required by new LIS professionals in the electronic world in the near future are concerned, we can identify different levels of skills. Firstly LIS professionals should have skills required for handling IT products, particularly, keyboard, operating system, softwares, physical handling of gadgets, tele communication products, DBMS, data and file management, DTP, word processing, generation of reports, etc. The next level skills include skills required to apply IT for service management in general and information processing, search and retrieval in particular. This involves collection and organization of data in electronic form, indexing techniques, selection and evaluation of sources, searching techniques, updating techniques, etc. Information retrieval skills include online searching as well as searching CD-ROM databases. This level should incorporate skills required for query formulations as well as query interpretation. The advanced level skills include internet skills and skills required for accessing networked resources as well as marketing of electronic information. A lot more can be said about skills expected for electronic publishing, electronic commerce and electronic marketing.

Hence advanced skills can be considered as a sort of specialization. Other IT related skills expected of new LIS professionals are skills required for evaluation of search results including modification of query for bettering the results and ranking of hit records and all that comes under post search processing and presentation of data/information. The post search processing is closely related to the personal information system (PIS) discussed earlier in the sense that customers should be trained to enable them to upload the data to the PIS. Hence PIS and post-search processing are also to be treated as enabling technologies and services. This aspect not only expects appropriate skills on LIS professionals but also presupposes ability to impart the same skills to customers.

4.8 Preservation and Current Use

Bringing usability and understandability into the discussion of preservation, with the implication that Representation Information needs to be collected,

has additional ramifications. In addition to the future users, who are the focus of preservation efforts of a digital object, there are current users who may also wish to understand and use those same digital objects. Current users will also require appropriate Representation Information. The tools and techniques which are needed for preservation are therefore also of use in making digital objects usable to a wider audience.

4.9 Challenges and Threats

We can consider some of the things can change over time and hence against which an archive must safeguard the digitally encoded information.

4.9.1 Hardware and Software Changes

Use of many digital objects relies on specific software and hardware, for example applications which run on specific versions of Microsoft Windows which in turn runs on Intel processors. Experience shows that while it may be possible to keep hardware and software available for some time after it has become obsolete, it is not a practical proposition into the indefinite future, however there are several projects and proposals which aim to emulate hardware systems and hence run software systems.

4.9.2 Environment Changes

These include changes to licences or copyright and changes to organisations, affecting the usability of digital objects. External information, ranging from name resolvers such as the DNS to DTDs and Schema, vital to the use and understandability, may also become unavailable.

4.9.3 Changes in What People Know

As described earlier the Knowledge Base of the Designated Community determines the amount of Representation Information which must be available. This Knowledge Base changes over time.

4.9.4 Termination of the Archive

Without permanent funding, any archive will, at some time, end. It is therefore possible for the bits to be lost, and much else besides, including the knowledge of the curators of the information encoded in those bits. Experience shows that much essential knowledge, such as the linkage bbetween holdings, operation of specialised hardware and software and links of data files to events recorded in system logs, is held by such curators but not encoded for exchange or preservation. Bearing these things in mind it is clear that any repository must be prepared to hand

over its holding – together with all these tacit pieces of information – to its successor(s).

4.9.5 Threats to Preservation

Surveys have been undertaken by PARSE. Insight Project and members of the Alliance for Permanent Access (http://www.alliancepermanentaccess.org/), investigating creation, re-use, preservation and publication of digital data. These surveys shows substantial demand for a science data infrastructure which is consistent across nations, continents and over a remarkably wide range of disciplines. There has been time for only an initial analysis of the results. The results of most immediate interest revolve around a collection of "threats" to digital preservation which are based on prior analyses of the domain and which are pertinent to data re-use also. It is worth noting that similar lists can be found in most project proposals related to digital preservation.
The major threats are as follows:

1. Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved
2. Non-maintainability of essential hardware, software or support environment may make the information inaccessible
3. The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity
4. Access and use restrictions may make it difficult to reuse data, or alternatively may not be respected in future
5. Loss of ability to identify the location of data
6. The current custodian of the data, whether an organization or project, may cease to exist at some point in the future
7. The ones we trust to look after the digital holdings may let us down

The preliminary survey results show that between 50% and 70% of responses indicate that all the threats are recognized as either "Important" or "Very Important", with a clear majority supporting the need for an international preservation infrastructure.
Another clear message is that researchers would like to (re-)use data from both their own and other disciplines and that this is likely to produce more and better science. However more than 50% report that they have wished to access digital research data gathered by other researchers which turned out to be unavailable.

Digital preservation is not a one-off activity; the variety of continuing threats and challenges, as outlined in this paper, must be addressed. The approach taken by OAIS has several benefits including providing the basis for certifying archives and showing how a preservation infrastructure is intimately connected with a general data infrastructure. This paper has expanded on these ideas and also described how the standard for audit and certification of digital repositories is being developed.

4.10 Conclusion

The preservation of the written heritage in whatever format it s being produced is of crucial significance to civilized society. Given that it is so important and that there are many strategic factors and costs which need to be established and predicted for the long term, it is an area where there are uncertainties. Digitization of cultural heritage materials is changing the ways in which collections are used and accessed. Many materials are amenable to digitization, including scarce, fragile and ephemeral materials, as well as the whole spectrum of moving image and audio materials. All can be safely used by wider audience in digital form. Research and interrogative tools for digitized source materials can also make digital surrogates more amenable to certain types of interpretation, such as full-text searching and indexing, as well as comparison of materials for multiple sources. Many factors will come into play when evaluating the value of digital resources, but these factors may help in assessing when digitizing collections can be cost effective. Valuable digital resources which will bring prestige to the institutions that create and maintain them will be those to support scholarship without any loss of the benefits of working with originals. When evaluating materials for digitization and evaluating whether or not the time is right for an institution to embark on such an initiative, it is important to consider the experimental nature of digital projects.

Much work that will be undertaken in the completion of digitization programmes will be at the bleeding edge of new technologies. This concept is relative: for institutions that have never worked with electronic resources before, all aspects of technology implementation can be traumatic. It will be necessary to ascertain the willingness and preparedness of institutions to embrace a certain degree of risk and experimentation and to understand whether or not such experimentation is acceptable or indeed necessary.

Developing understanding of the information that will be needed to preserve digital content has been an impressive co-operative effort on an international scale. It is an ongoing task and technology is not standing still while we figure it out. However great gains have come and will continue to come from sharing expertise throughout the digital archiving community. Cooperative efforts including metadata, standards, metadata extraction and conversion tools and format registries offer the hope of a longer life for digital content worldwide.

Self-assessment Questions

1. Discuss the purposes and advantages of digital preservation.
2. How various technologies are helpful in digital preservation of library material?
3. Describe various reference and information services provided in digital environment.
4. Explain various challenges and threats faced by libraries in preserving the print material in digital format.

Activity:

1. Visit the website of National Archives of Pakistan (NAP) and prepare a report on major functions of NAP in respect to preservation, microfilming and digitization and restoration of archival material

Recommended reading:

1. Caplan, P. (2008). *The Preservation Of Digital Materials; Expert Guides To Library Systems And Services*. Chicago, IL: American Library Association.

2. Council on Library and Information Resource. (2005). *CLIR Publication No. 129*. Deegan, M., & Tanner, S. (Eds.), *Digital Preservation*. London: Facet Publishing.

3. Glister, P. (1997). *Digital Literacy*. Hoboken, NJ: John Wiley And Sons Inc.

4. Gorman, M. (2003). *Enduring Library*. Chicago, IL: American Library Association.

5. Hughes, L. M. (2004). *Digitizing Collections. Strategic Issues for The Information Manager*. London: Facet Publishing.

6. Lankes, D. R. (Eds.). (2000). *Digital Reference Services in the New Millennium; Planning Management and Evaluation*. New York: Neal Schuman Publishers Inc.

7. Mcclure, C. R. (2002). *Statistics, Measures And Quality Standards For Assessing Digital Reference Library Services: Guidelines And Procedures*. Syracuse, NY and Tallahassee, FL: Syracuse University and Florida State University.

8. Pace, A. K. (2003). *The Ultimate Digital Library; Where the New Information Players Meet*. Chicago, IL: American Library Association.

Unit.5

DEVELOPING COLLABORATIVE DIGITAL LIBRARY NETWORK AND INTELLECTUAL PROPERTY RIGHTS

Complied by: Muhammad Jawwad

Reviewed by: Dr. Amjid Khan

CONTENTS

INTRODUCTION

The unit is designed to explain Collaborative Digital Library Development Networks and some of the prevailing social networking issues affecting digital library initiatives, particularly the collaboration patterns among participating institutions as well as funding agencies. It also tries to identify social relationships amongst the networked institutions in terms of nodes and ties. The unit will also critically appraise the formation of a formal social network in the digital library initiative project embracing local memory institutions. Moreover, the unit will focus on the advantages of digitising information. But will highlight a serious issue that has cropped up due to it is the unethical use of information. Intellectual Property Rights (IPR) has come up as one of the most important and debated issues due to digital information. The unit discusses all the issues in detail.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is collaborative digital library network?
2. What is meant for collaboration model and digital libraries?
3. How can libraries start and develop collaboration in the digital age?
4. What is intellectual property right?
5. How intellectual property rights changing the libraries digitization projects?
6. What are categories of intellectual property rights?
7. How important are intellectual property rights in digital era?

5.1Introduction

Collaboration can be defined as a formal set of relationships between individuals and organizations in a development process that exists through the cycle of idea formulation, design, product testing and implementation. This form of collaboration relies on a significant input of human energy, resources and commitment over a sustained period of time. The range of possible collaborative ventures encompasses one product/one-time partnerships to complex relationships that develop successive products or technologies for years.

The rapid evolution of new information technologies presents fertile opportunities for collaborative efforts in higher education. Collaboration has become an important concept in technological development primarily because of its potential to better leverage resources, including access to critical hardware and software products. Other advantages of collaborative efforts include the minimization of risks, the encouragement of innovative thinking, shared expertise, expanded access by users and improved productivity.

Sectors of the economy under rapid change, such as information technology, require collaborative efforts to develop effective hardware and software solutions because of the significant investment in technical design and problem-solving for a short product life cycle. An essential advantage in partnering is the ability to cut costs and improve outcomes for each organization by allowing broader (aggregated) resource inputs in the design through implementation stages of a project.

5.2 Collaboration model and Digital Libraries

The collaborative model is becoming fairly prevalent in the development of digital library resources. Cornell University has several electronic library projects that have been generated in the last four years. In an issue in *Library Hi Tech* in 1994, the Cornell University Library System highlighted eight digital library projects, all of which are collaborative in nature. Many involve industrial partners including Kodak and Xerox, and publishers such as Elsevier and UMI. Several involve cooperative ventures with other research institutions, foundations and government agencies. It is clear that the scale of large digital projects require testing on a broad scale, a critical mass of documents to digitize, remote access opportunities and

financial support and most importantly shared intellectual resources for ideas, development and problem-solving.

Another important digital library project that features extensive collaboration is the JSTOR initiative. This large journal digitization project began in 1993 through the auspices of the Mellon Foundation, involved Princeton University in the idea generation stage, Harvard University provided the hard copy journals for production and the University of Michigan serves as the host computer storage facility for Internet distribution of the digitized articles. Initially, a pilot group of five colleges, including Bryn Mawr, Denison, Haverford, Swarthmore, and Williams tested the use of the electronic resource among faculty and students. Now the benefits of this collaborative effort can be shared more widely as the JSTOR digital journals are being marketed as a digital library component for other libraries.

An associated benefit also accrues from the collaborative process: participation in the design enhances the potential for acceptance of the solution and a willingness to implement it. A study in the *Sloan Management Review* on successful information system building in industry suggests that concurrent system development that includes collaboration between engineers, developers and implementors will help insure information system implement ability primarily because internally integrated and consistent ideas emerge from the collaborative process that are easy for users to accept because they have been communicated, evaluated, compared and debated holistic holistically.

We have modeled several of the critical components of the digital library: course reserves, digital journals (that is print journals that are digitized), archives, images, CD-ROM databases (non-index, non-journal) and media files, that is audio and video. The text files are relatively easy to manage and support and scale with predictable results. Images, including photographs, art reproductions, graphs and maps are also manageable. However, audio and video files present several significant problems: from standards for access and delivery to storage concerns for scaling to major rights management problems.

The building of a digital library is an event that must be conceived in extended time periods. Most private liberal arts college have been building their library print resources for at least one hundred years. A digital library will take a considerably

shorter period of time, since many large databases can be purchased commercially taking advantage of digital economies of scale in distribution, storage and access. However, the digitization of original local holdings--that is faculty and student course work and special collections-- must be viewed as a re-publishing effort that takes considerable time in planning, selection, scanning, loading, indexing and intellectual property rights management, but will pay dividends for the college curriculum as course material becomes available longitudinally and in more depth.

A true digital library will have to include locally unique materials produced by faculty and students and library archival materials that no other institution holds. These materials must be made searchable locally and remotely. The library will become publisher of student and faculty scholarly works that have not been previously collected or generally accessible. Thus the original impact of the OCLC Union Catalog on ILL in opening small library collections and relieving the burden on research collections, will be repeated when colleges mount digital collections of unique holdings on the Internet.

The scaling effort to build a digital library rapidly incurs many considerations: number of items to be stored and accessed, the number of databases to be supported, the number of access points to be designed or enabled, the number of users to be supported, the level of functionality to be implemented and the size of the digital items relative to bandwidth capability. Some of these issues are management problems, some are resource questions, and several are technical considerations that must be taken into account. Together they represent a complex problem-solving opportunity that must be carefully planned and executed.

5.2.1 Partners are Available to Assist in the Development.

Right now many people are excited about the possibilities generated by building digital libraries. Take advantage of the opportunities to work with faculty and students on campus first. They are the source of the best ideas, which will enable you to avoid some of the pitfalls of building librarian-centered systems. Look to collaborate with technical staff on campus next. Buy them dinner and drinks as often as feasible. Seek information industry partnerships. Many vendors are looking for colleges to work with. Be inclusive as much as possible, there are a lot of possibilities out there. Finally, the best partners are the academic libraries in your region, get together and start planning how you can work together to reduce costs, build content and increase access.

It seems likely that the best of the next generation of digital library initiatives will be founded on deep levels of collaboration across a broad spectrum of parties and contributors, including scholars, librarians, technologists, press directors and other publishers, scholarly societies, and ordinary citizens interested in developing new information resources dedicated to and available to the population at large. But such levels of collaboration have been difficult to initiate, and difficult to sustain— perhaps because they have been developed with existing systems and services as the baseline. We have the opportunity, now, while a window is open, to free ourselves from the rigid boundaries of these systems, develop new methods for working across boundaries and disciplines, and realize the promise of technologies that could remove barriers of time and place— if we only let them. In short, what is holding us back from a quantum leap in development of large-scale digital libraries is not technology, but the existence of long-established systems and structures for carrying out the discreet elements that are necessary in the creation of a broadly conceived digital library— one that is not only built upon commercially-licensed information resources but encompasses entirely new resources that have no parallel in the print or commercial sector.

5.2.2 Libraries and Collaboration

Cooperation, collaboration, and resource sharing are much of what makes the library enterprise so successful at meeting user needs. Libraries led the way in terms of collaboration in higher education before it became both fashionable and necessary to do so. A cursory review of the literature of collaboration in higher education shows that library collaboration is often the first, the most long-lived, and the most successful of collaborative endeavours among a group of universities. These arguments can be extended that other realms of the university such as information technology, the university press, and the academic and research enterprises of the modern universities. In each case, systems have been developed— some automated, some not, but all intended to streamline and organize the administration of these divisions— and it is difficult to effect change while working within such systems.

Let us take a look at two areas of traditional library collaboration as examples of the kinds of systems that are in place. Interlibrary lending and resource sharing were not part of the original canon of library operations. These services arose in response to the availability of shared bibliographic records and concomitant delivery systems that could support them. And, because these systems and services are relatively

modern, and grew from individual efforts, they have become standardized only in recent times. Nevertheless, a set of processes comprising the interlibrary lending "system" developed relatively quickly, and those systems have remained relatively unchanged. To be sure, aspects of the system have been automated, but true innovations have been long in coming.

Contemplate for a moment the many person-hours devoted to managing routine interlibrary lending tasks in any library today, in spite of the availability and adoption of large-scale national utilities for this purpose. It should be obvious that certain functions related to the service aspect of the enterprise (e.g., how the patrons request information, manage their accounts, and receive the materials) could be centralized and routinized, and others eliminated to achieve efficiencies in operation and enhancements in service, but few libraries have radically altered their systems or services in this area (a notable exception being Colorado State University, which completely revamped ILL after a catastrophic flood destroyed their records and systems). Indeed, many have persisted in operating systems that are more tuned to the 1960s than the 21st century.

5.2.3 Libraries and Cooperative Collection Management

Let us turn our attention to cooperative collection management, which with a few notable exceptions has simply languished with its promises unrealized. It is obvious that no library can ever hope to collect everything published in every discipline, yet few libraries have successfully initiated programs and plans for coordinated, cooperative collection management— even in those instances where libraries are in close physical proximity and share a user base. The clear exception to this rule is in the realm of collaborative acquisition of commercial electronic texts and/or the collaborative licensing of electronic resources. And it is precisely because "systems" for acquiring and managing digital or electronic content have yet to emerge that innovation is flourishing in this area.

Within the digital realm, where development of large-scale initiatives are yet in a state of relative infancy, we have before us the opportunity to establish processes and services that take full advantage of this medium in order to better serve society at large. As we go about the necessary and desirable work of building many separate and distinctive digital collections, we can base our developments on the assumption

that these materials will be shared rather than waiting until a fully developed (and potentially inflexible) "system" is in place.

In short, new inter-institutional request, delivery, and collection management have been difficult to implement because the systems in place for managing these aspects of the libraries have been in place for many years. It has been possible to cooperate around the margins, or to innovate at the edges of these systems, but librarians themselves have clung to the systems at the expense of innovation that would revolutionize service and access. We have the opportunity to carefully consider the "systems" that we put in place for sharing the development of and access to digital information sources. Now is the time to seriously consider this— lest this brief window for innovation close and an inflexible system emerges of whole cloth— widely adopted and forever calcified. For this purpose, I will focus on the attributes of collaboration as it might relate to emerging digital efforts.

5.2.4 Elements of Good Collaboration

Well-managed collaboration allows innovation, creativity, and expansiveness. Undisciplined, unfocused collaboration can reinforce the status quo, block innovation, and divert attention from the core business of any organization. Successful collaboration is marked by the establishment of a framework within which many and varied activities may arise. In this context, collaboration may evolve and develop based on the needs and evolutionary development of the collections and the users they are intended to serve.

Before fully examining the elements of successful collaborations— those elements we may wish to consider and replicate in the digital environment— we might first examine the hallmarks of poor collaboration. To mangle Tolstoy, it might be said that all bad collaborations are bad in their own way, and yet they might be divided into two "types." The first of these types include cooperative efforts that start out with the best intentions, but devolve into peer-driven associations with no clear overarching strategic aim— or worse, competing aims. These collaborations suffer from "goal drift" and lack of support at the highest levels of the organization. The second type of ineffective collaborations are those that are established within the organization by well-meaning individuals who, nevertheless, fail to achieve buy-in from the administration. Such efforts are destined to remain "boutique" activities which may well benefit the

individual participants (particularly in terms of personal development and professional growth), but rarely have a positive impact on the organization.

Good collaborations, by contrast, share following common elements:

- A clear leader or leaders— in positions of authority within and across the participating partners— who advocate for the collaboration (collaboration is, after all, hard work);
- Active, engaged participation by those who have the ultimate budgetary and management oversight of those activities being addressed;
- An identified and commonly understood (if not codified) strategy— a clear understanding of the purpose and benefit of working together; and
- Constant and clear communication with all parties involved, demonstrating benefits of the enterprise, seeking feedback, and adjusting.

We can suspect that those who malign collaboration are actually frustrated with the ways in which cooperation is managed and manifested. It is possible, of course, to engage in bad collaboration, just as it is possible for any single person to wrongly go about his or her own work, or for an entire department of staff members to delude themselves into thinking they are doing a "good job." And, just as it would be wrong to gauge an entire staff on the actions of one who fails miserably at his or her job, so it seems wrong to gauge all collaborations as ineffective.

5.2.5 Attributes of Digital Publishing and Collaboration

Like any endeavour involving humans with varied interests, goals, and objectives, the collaborative enterprise will run into the tensions inherent in such activity. The primary tensions include the perception (if not the reality) that the collaborative effort will limit autonomy and independence of the participants, and will slow the decision-making process. Both tensions can be successfully addressed, of course. A true collaborative activity will always remain flexible and voluntary— allowing participants to invest at varying levels as warranted. And decision-making structures can be implemented that allow for broad input without dramatically slowing decision making.

To these issues, we might add a number of attributes that we might strive to replicate in building new processes for digital publishing and content management. For example, the processes identified:

- Provide a framework that allows individual and collaborative development;
- Are simple enough to encourage adoption;
- Do not compete with local interests and obligations;
- Enable cooperation without mandating it; and
- Engage many at the local level, including creators, users, and organizers of information.

Turning to specific tasks that might profitably be addressed in such a collaborative, cross-disciplinary environment, experts might be brought together to address issues of content creation and preparation, editing, peer review, acquisition of external or commercial resources, access, archiving and preservation for perpetuity. Ideally, these collaborative discussions would develop within the parameters identified (e.g., through the establishment of a simple framework that does not compete with local interests and which enables cooperation at many levels). In turn, these discussions would result in a set of national or international interlocking agreements or guidelines that establish both the technical and organizational guidelines for the development of digital collections. It is assumed that such agreements would maximize opportunities for sharing and minimize opportunities for commercial manipulation of the resulting content, collections, or access.

Note the absence in the above paragraph of any mention of a "business plan" or "economic model." To be sure, no enterprise can move forward without resources, be they fiscal, human, or infrastructure, but libraries and universities are not "businesses." It is precisely because we exist outside the normal boundaries of profit and loss that we can be creative in devoting time and resources to these efforts— and we should. The question is not "how much should we charge others to use this content?" but rather "what is this effort worth to this organization?" Just as universities may be perceived as major contributors to the economic development of a community or state, should not we keep in mind that intellectual development is just as critical to a healthy society? And shouldn't experiments in the capture and delivery of new forms of intellectual "products" be as important as the "research parks" so fashionable on university campuses?

5.3 Intellectual Property Rights and Digitization

The path-breaking developments in information technology (IT), particularly in the field of computers, communications and mass storage, have made it possible to

handle enormous volumes of digital information and data with ease. CD-ROM has become an acceptable medium for storage and dissemination. More and more academic, R&D and other institutions are turning to CD-ROM databases replacing some of the conventional publications and online databases slowly but steadily. The number of electronic databases of secondary periodicals is also growing. Retrospective and bibliographic searches and reference queries are exclusively conducted using them. Digital information would have a higher growth in the coming years. Library and information centres (L&Ics) are taking advantage of these technologies for effectively meeting user's requirements. There is a gradual increase in the digital information component in the holdings with a growing dependency on electronic resources, particularly in S&T and academic libraries. During the past few years, there was a computers; the focus of the virtual library being the individual users or their workstations. In a true virtual library there is no corresponding physical collection, documents are available in electronic format, the) are not stored in any one location, they can be accessed from any workstation and are retrieved and delivered as and when required, and effective search and browse facilities are available. The usage of communication networks for access, browsing, downloading or transmitting information is essential in a virtual library environment.

Every human being is endowed with certain but varying degree of intellect. Each individual is uniquely gifted. The word intellect originates from the root “intellectus” in Latin which means the power of knowing as distinguished from the power to feel. Man has the capacity to acquire knowledge and increase his knowledge bank by gathering more and utilizing it as and when required throughout his life time. An intellectual makes his living by selling the product intellect, which is nothing but the brain child of his original idea, creative thought, which forms a special kind of property known as intellectual property. The intellectual property is ownership of something intangible. A right as we know is legally protected interest and object of the right is the thing in which the owner has this interest. The object in intellectual property right is immaterial property. Hence, we cannot neatly fit it into either of the two conventionally dichotomized categories of personal or real property. Intellectual property (IP) refers to creations of the mind: inventions, literary and artistic works, and symbols, names, images, and designs used in commerce

5.4 Categories of Intellectual Property Rights

Advances in digital information have changed drastically the scenario of information access. It enables anyone to produce digital copies of works that are as good as original. Parts of works can be easily copied to create new documents due

to digital technology. An individual's efforts behind creating something need to be protected. Law has come to the rescue of creators of knowledge. It protects their creations in recorded form. Such rights are termed as Intellectual Property Rights (IPR), provided to the creators/inventors of intellectual works. IP is divided into following two categories:

5.4.1 Industrial Property, which includes inventions (patents), trademarks, industrial designs, and geographic indications of source; and

5.4.2 Copyright, which includes literary and artistic works such as novels, poems and plays, films, musical works, artistic works such as drawings, paintings, photographs and sculptures, and architectural designs. Rights related to copyright include those of performing artists in their performances, producers of phonograms in their recordings, and those of broadcasters in their radio and television programs.

IP as a general term refers to the subject matter of the laws that give rise to proprietary interests in creations of the mind. That is, it refers to intangibles that arise and derive their intrinsic value out of innovative or creative activities. The various tools of IPR that are used to protect innovations are Copyrights, Industrial Designs, Data Protection, Geographical Indications, Patent and Trademark.

As discussed, IPR are divided into copyright, patents, service and trade marks and design rights. It includes copyright for literary, graphical, musical and other artistic works; patents, trademarks, layout designs of integrated circuits. IPR comprise moral rights and economic rights. Moral rights help preserve the integrity of the work and attribution of the work to the author. Economic rights provide economic benefits to the creator.

The scope of IPR as the right to: reproduce/copy, prepare derivative works, distribute its copies publicly, do public performance and public display. IPR also limits rights of libraries to reproduce, loan out to other libraries and also of individuals to make limited copies for use or archiving in the context of fair use.

5.4.3 Patent

Patent is granted for an invention. It may be a process showing a new way of doing something or for a product created that is new. Novelty is important in the product or process for the patent to be granted. The criteria for an invention to be granted a patent is:

- It should be novel.
- It should have inventive step or it must be non-obvious.

- It should be capable of industrial application.
- It should not fall within the provisions of section 3 and 4 of the Patents Act 1970".

Certain subjects are not patentable e.g., methods for medical treatment, scientific theories or discoveries of natural elements. Patents are beneficial to the society as they promote creativity and innovation. Patents have a long history. Edison got a patent for electric bulb. "Patent protection for inventions is older than copyright. The earliest patents giving the exclusive right to exploit an invention were awarded in Italy in the 1400s. In England the Crown awarded patents giving inventors a monopoly to exploit their inventions in the same way it gave printers monopolies to print books."

5.4.4 Trademark

Trademark is a "word, phrase, symbol or design, or a combination of words, phrases, symbols or designs, that identifies or distinguishes the source of the goods of one party from those of others" (United States. Patent and Trademark Office). A more elaborate definition of trademark "is a visual symbol which may be a word signature, name, device, label, numerals or combination of colours used by one undertaking on goods or services or other articles of commerce to distinguish it from other similar goods or services originating from a different undertaking" (India. Controller General of Patents, Designs and Trademarks. Office of the Registrar of Trademarks).

5.4.5 Design

Design "means only the features of shape, configuration, pattern or ornament or composition of lines or colour or combination thereof applied to any article whether two dimensional or three dimensional or in both forms, by any industrial process or means, whether manual, mechanical or chemical, separate or combined, which in the finished article appeal to and are judged solely by the eye, but does not include any mode or principle or construction or any thing which is in substance a mere mechanical device, and does not include any trade mark, as define in clause (v) of sub-section of Section 2 of the Trade and Merchandise Marks Act, 1958, property mark or artistic works as defined under Section 2(c) of the Copyright Act, 1957." In India it is registered under the Designs Act 2000 and granted for a period of 10 years that can be extended for 5 years more on application. (India. Controller General of Patents, Designs and Trademarks. Department of Industrial Policy and Promotions. Ministry of Commerce and Industry).

5.4.6 Copyright

Copyright gives the creator, the right to protect his/ her creation. Protection does not

mean that access to the creation is forbidden. If it so happens, it will hamper further growth of knowledge since creation of knowledge requires information as an ingredient. It is important to note that a copyright exists in form and not in idea. For instance, if a person were to write a classical love story based on the idea, in the settings of the early 18th century, it is not that nobody can use the said idea and write another novel, but it is merely, that the novel of the first writer should not be copied.

Thus, copyright is assigned to documents or recorded work and not to ideas. It can be assigned to individuals or to organizations. An example of an organization earning a copyright is when its employee works and produces a document pertaining to its policies, objectives, functioning, etc. An author may transfer copyright to his/her publisher for reproducing copies of the work. But that does not imply that the ownership of the work is transferred to the publisher. In this case the author has transferred exploitation/economic rights to the publisher but the moral rights remain with him/her. He/she has all rights to be acknowledged for the work and not to permit changes and mutilation to the work. Any copyrighted work should have:

- A means of identifying each work and its copyright status;
- A means for assuring that conversion via scanning results in a digital form that includes the identity of the work, the copyright status and whether permission for a conversion has been granted;
- A means for authenticating each work;
- A means for protecting each work so that only an authorized recipient can receive and perceive it;
- A means for controlling, and setting limits on, specific uses (e.g., display only, print one copy, and no downloading);
- A means for right protection of right protection for each work so it cannot be altered;
- A means for metering usage;
- A means for electronic contracting for access and use;
- A means for billing and collecting payment;
- A means for establishing the perimeters of authority for software agents, including, for example, limiting the amount that can be obligated in electronic contracting conducted by software agents; and
- A means for assuring that copyright identification and some means of control stay with every portion of the work when it is downloaded or otherwise transferred in digital form, or when it is printed."

5.5 Importance of Intellectual Property Rights

Importance of IPR The accelerating pace of technological improvements _ some

radical and mostly incremental has resulted in fiercely competitive markets for all types of products and services, including machine tools. For any enterprise to come into existence, survive and thrive in such competitive arena requires access not only to the traditional factors of production (land, labour and capital) but also to cutting edge knowledge in a diverse range of relevant fields technical, financial, legal and commercial. In relation to any needed inputs for business, every organisation has to decide whether to develop new or improved technologies on its own, in partnership with others or buy / license it from others, if possible.

For instance, the desired output of R&D may be obtained through in-house R&D, through joint R&D efforts in partnerships with a range of partners (such as, other businesses, research institutes or universities) based in Pakistan or overseas. This could be through subcontracting or outsourcing agreements with vendors, suppliers, etc., and may be licensed or bought of-the-shelf, if available, in the marketplace.

Similarly, designing and / or marketing of a new or improved product, may be done either independently under an organisation's brand-name (trademark), or may be left to other companies, which may use the organisation's manufacturing capability to execute their technical specifications (based on their inventions, innovations and technologies). The other company can take the product, to market that was manufactured by the organisation as a contract manufacturer or OEM supplier, under the other company's brand (trademark). These kinds of possibilities would be equally true for machine tool companies in India or elsewhere For a machine tool manufacturer, irrespective of its standing in the domestic or international value chain/network, the only way to climb up the ladder of profitability and success is through differentiation of its products and/or of its organisation from its competitors.

This differentiation could be brought about by superior ability to create, access, protect, exploit and manage efficiently and effectively, in domestic and export markets, the knowledge-driven competencies. These could be done through proper use of a whole range of tools provided by the system of IPR. Faster pace of innovations and their strategic use and effective management is critical for developing an internationally competitive business in any innovation-driven industry, including machine tools. This requires companies in this industry to develop, amongst other things, ability to create, identify, protect, use, share and manage self-created or acquired IPR assets as an integral part of their business strategies so as to realize maximum value out of their inputs, processes and outputs.

The basic tools for the protection of IPR are as follows: • Trade Secrets • Patents • Industrial Designs Registration • Trademarks (including service marks) • Copyright • Geographical Indications • Layout Designs for Integrated Circuits.

5.6 Copyright and Access Educational Materials

In recent years, there has been a welcome expansion of primary and secondary education in developing countries, and aid has been concentrated rightly in these sectors. Whilst there are still major challenges in achieving "Education for All", developing countries and their donor partners are making significant progress. Access to books and reading materials at primary and secondary levels in some countries has also improved. This is the result of increased levels of public expenditure on primary education and international book donation programmes, such as Book Aid International. And importantly, in some countries, it is also because local publishing industries, albeit often at an embryonic stage, are able to produce low cost schoolbooks and reading materials. However, access to books and learning materials is still a real problem in many developing countries.

Developing countries need educated people such as doctors, nurses, lawyers, scientists, researchers, engineers, economists, teachers and accountants. Without people skilled in these professions and a system of life-long learning and education, developing countries will be less able to absorb new technologies, generate innovation, and compete in the global knowledge economy. For example, even if developing countries can obtain cheap medicines they will still need trained doctors and nurses to administer them properly in order to save lives. With many developing countries already spending a significant proportion of GNP on education, they may be unable to find the additional resources required simply to maintain current levels of tertiary enrolments, let alone improve quality. Clearly, copyright is not the only issue with the weak tertiary infrastructure but high prices of books and materials and limited access to Internet-based resources are still important parts of a worsening crisis. In the tertiary sector, the evidence indicates that access to books and other materials for education and research remains a critical problem in many developing countries, particularly the poorest. Most developing countries remain heavily dependent on imported textbooks and reference books, as this sector is often not commercially feasible for struggling local publishers to enter. The prices of such books are beyond the means of most students.

University libraries should play a key role in supporting research and ensuring access to copyrighted books, journals and on-line materials for poor students in developing countries, but they are typically in a very poor state. Donor agencies have provided funding to modernise and re-stock libraries in a number of countries, including providing Internet connectivity and photocopying facilities. Indeed, even well resourced libraries in developed countries are experiencing extreme difficulty in continuing to stock the full range of journals their academics and students expect. In developed countries the rapid increase in subscription prices for academic journals, and ongoing consolidation in the publishing industry has fuelled an active debate on how researchers can maintain access to the materials they need, and the development of alternative models of on-line publishing. But developing countries also need to be allowed greater freedom to relax international copyright rules to meet their educational and research needs. As we have noted, delegates at the Stockholm conference proposed a package of such amendments to the Berne Convention in 1967. Developed countries rejected these proposals because they were considered to place too radical limitations on copyright protection.

Examining the evidence 30 years later, it is clear to us that the special provisions for developing countries that were added to the Berne Convention in 1971, as set out in the Appendix, have not been effective. Further reforms are therefore needed, and different measures may be more or less important in meeting the specific needs in individual countries. As one commentator has put it: “In some cases, access to scientific journals and books at subsidized prices for a limited period would help greatly. In others, local publishers with limited markets need easy and inexpensive access to foreign books in order to translate them into the local language. In a different context, permission to reprint books from the industrialized countries in the original language is needed to serve an indigenous population literate in English or French but unable to pay the high cost of imported books. And for some countries, most of the elements of an indigenous publishing industry are missing and there is a need to build it up from scratch. Copyright may not be the key element in all of these circumstances, but it does play a role.”

To improve access to copyrighted works and achieve their goals for education and knowledge transfer, developing countries should adopt pro-competitive measures under copyright laws. Developing countries should be allowed to maintain or adopt broad exemptions for educational, research and library uses in their national copyright laws. The implementation of international copyright standards in the

developing world must be undertaken with a proper appreciation of the continuing high level of need for improving the availability of these products, and their crucial importance for social and economic development.

5.7 Intellectual Property Rights in Digital Era

IPR issues in the IT sector have assumed significance following rapid emergence of digital era. IPR cells are food to be set up in different countries of the world to launch awareness programmes in collaboration with Public Sector undertakings, industries, scientific societies, etc. as in India and Pakistan. Such cells extend support to file applications for patents and copyrights. Recent advances in technology have produced radical shifts in the ability to reproduce, distribute, control and publish information. WWW has allowed everyone to be a publisher with worldwide reach.

At this crucial scenario, in order to protect the right of authors, developers and to provide privacy and liberty to user, it becomes essential to introduce laws against computer crime and help to build IT infrastructure. In view of these facts, various laws, Electronic-fund transfer, Electronic cash digital intellectual property rights, etc. have been enacted. In India and Pakistan specific studies are undertaken to enhance the understanding of IPR issues in the field of Software, Multimedia, National Information infrastructure, global digital economy, cyber laws, and feasibility of electronic right management system in the line with WIPO treaties on digital agenda. The initiatives taken by the Government of India, Private and NGOs to create awareness about IPR are creating a climate to face the challenges of the new IPR regime.

Chinese copyright protection system has raised specific issues of IPR in digital context. These are also relevant in the Pakistani context. The advent of digital technology has greatly accelerated the dissemination and distribution of information with great speed and accuracy never seen before. It is much easier to disseminate literary, artistic and scientific work to a very large community of Internet users and users of electronic media. At the same time poses some problems and issues for consideration. The major issues are,

- Is digitization to be considered as similar to reproduction, for example using Xerox machine?

- Is digitization a deductive activity such as translation from one language to another?
- Can transmission of digitized documents through Internet be considered as Commercial distribution or public communication similar to broadcasting?
- Is the principle of exhaustion of the distribution right still effective in the digital age?
- Can we consider a database as a special collected work that should be protected by the copyright law or it can be considered as a special work requiring specific legislation for its protection?
- What can be considered as "Faire use" in the Internet environment? • What are the concerns of the library community?
- In the digital context if access could be technologically restricted by the copyright owner, how could the public exercise fair use with regard to those work?
- Whether libraries should be prevented from employing digital technology to preserve work by making three copies-an archival copy, a master copy and a use copy?
- Whether Internet Service Providers (including libraries and educational institutions) should be liable for copyright infringement merely because they facilitated the transmission of digital data (Zeroes and Ones) that translated into another party's copyrighted work.

The issues mentioned above are specific to the library community. The libraries as a Service have allowed their users to read a document, to browse through the whole collection; to search through the library catalogue; to supply Xerox copy for specific individual research and education purpose; to procure photocopies of articles from other libraries or clearing centers; to widely distribute the re-produced copies of documents requiring public awareness and to provide inter library loan service. Whether all these activities will continue in the digital age? If digitization is considered as reproduction, it is clear that in digitization the initial work is merely changed into the digital form and the Process of changing is accomplished by a machine, without any creativity.

At the same time if it is considered as a translation from one language to another, the digitization is also a change from natural language of humans in to binary language of machine. In digitization however, there is no creativity involved and it could be considered as an activity similar to reprography. The copyright protects

creative works. Simply transformation in to the digital form of an original document cannot be considered as creative. The transmission of information on Internet can be considered similar to broad casting and copyright law cannot be applied. Internet transmission is global in nature. A tangible object carrying a traditional work distributed lawfully comes under the principle of exhaustion of distribution right. It is not appropriate to apply this right to a work transmitted on Internet. The transmission on Internet is different from any tangible object fixing of the work.

5.8 Conclusion

A database is normally defined as the structured storage of data in a computer system. As the selection of data and arrangement of data in a structured way could be considered as intellectual creation of database, accordingly, should be protected by copyright law. The fair use of print material by allowing reproduction in a reasonable way for private study, research or education is well understood. But in the context of digital information, because it is distributed to a larger community, it is difficult to judge, comprehend "fair use", access and control the infringement of copyright law. It is almost impossible for a copyright owner to know which person used his/her work. It is also impossible for copyright owner to give permission to use and receive remuneration. In this context it is necessary to modify the copyright law. The librarians in the digital environment have the same responsibility to collect information and help the readers by giving it even if the form is electronic information. The role of librarian is to be protected and enhanced. The copyright protection should be encouraging the use of information for creativity and not for creating hurdles in the use of information. The Librarians should continue to work as catalyst for the free flow of information between the owners of copyright and the users of the information. Collaborations are not immune from the vicissitudes of the typical academic environment or corporate or public bureaucracy. However, when effective, they can create appropriate pressures for change and/or eliminate barriers that might exist. In the complex world of digital content development, in a world where researchers, scholars, technologists, and librarians must work outside of existing systems to achieve their goals, collaboration will be an essential element of success. In so doing, we will lead the way to more effective and creative means of accomplishing our goals— and we will do so in an environment that supports and responds to the intellectual and the common good— not to the marketplace.

Librarians are experienced collaborators, particularly with each other. Librarians across the country copy catalog records, expand local services through interlibrary loan, develop best practices for services through professional associations, and work together on advocacy for information law and policy. Librarians also collaborate to create digital collections with a goal of broad fee-free access to information, sometimes with other libraries and sometimes with commercial partners. The future of libraries will probably build on existing collaborations in order to pool resources for common goals, and because access to digital collections reduces the need for library users to make a trip to a particular library location. The power of collaboration promises a much richer pool of materials than any one library can obtain and curate. Research libraries have already contracted with Google to provide content for the Google Books project, while also developing the library preservation consortium known as HathiTrust for library-controlled access to copies of these digital records and other digital collections.3 Visionary librarians have helped create the Digital Public Library of America (DPLA) to increase access to information in libraries of all types from across the country. Librarians have contributed to the movement for open access to scholarship and have shaped related efforts like Harvard Law Library's Caselaw Access Project etc.

Self-assessment Questions

1. Discuss collaborative digital library networks and various models of collaboration.
2. Why should libraries consider the intellectual property rights while starting digitization project?
3. Explain various categories of intellectual property rights with examples.
4. Discuss the importance of intellectual property rights in present digital era.

Activity:

1. Visit the Internet and explain the historical background of JSTOR digital collaborative library network along with various other universities library networks.

Recommended Reading:

1. Agnew, G. (2008). *Digital Rights Management: A Librarian's Guide to Technology and Practice*. Oxford: Chandos. Ahuja, A. (1996). Chaining the Unchained Books: Copyright as an Infringement on the Philosophy of Library Science. *DESIDOC Bulletin of Information Technology*, *16*(6), 5–10.
2. Alikhan, S., & Mashelkar, R. (2004). *Intellectual Property Rights and Competetive Strategies in the 21*st *Strategies*. The Hague: Kluwer Law International.

3. Ananda, T. B., & Reddy, R. B. (1996). Copyright in the Information Age: Librarian's Viewpoint. *DESIDOC Bulletin of Information Technology*, *16*(6), 17–19.

4. Balakrishnan, N. (2005). Universal Digital Library – Future Research Directions. *Journal of Zhejiang University. Science*, *6A*(11), 1204–1205. doi:10.1631/jzus.2005.A1204

5. Balakrishnan, N., Reddy, R., Ganapathiraju, M., & Ambati, V. (2006). Digital Library of India: a testbed for Indian language research. *TCDL Bulletin, 3*(1).
6. Caslon Analytics. (2007 February). *Moral Rights. Caslon Analytics Intellectual Property*. Braddon, ACT, Australia: Caslon Analytics. Retrieved from http://www.caslon.com.au/ipguide18.htm

7. Controller General of Patents. Designs and Trademarks. (n.d.). *Frequently Asked Questions*. Department of Industrial Policy and Promotions, Ministry of Commerce and Industry. Retrieved from http://www.patentoffice.nic.in/

8. Controller General of Patents. Designs and Trademarks. (n.d.). *Frequently Asked Questions*. Office of the Registrar of Trademarks. Retrieved from http://ipindia.nic.in/tmr_new/default.htm

9. Das, A. K. (2008). *Open Access to Knowledge and Information: Scholarly Literature and Digital Library Initiatives – the South Asian Scenario*. New Delhi: UNESCO. Retrieved February 1, 2009, from http://dlist.sir.arizona.edu/2281/01/ Open%5FAccess%5FBook.pdf

10. Das, A. K., Dutta, C., & Sen, B. K. (2007). Information Retrieval Features in Indian Digital Libraries: A Critical Appraisal. *OCLC Systems & Services: International Digital Library Perspectives*, *23*(1), 92–104. doi:10.1108/10650750710720793

11. Hofman, J. (2009). *Introducing Copyright: A Plain Language Guide to Copyright in the 21*st *Century*. Vancouver, Canada: Commonwealth of Learning.
12. Koelman, K. J., & Helberger, N. (2000). Protection of Technological Matters. In *Copyright and Electronic Commerce: Legal Aspects of Electronic Copyright Management* (pp. 165–227). The Hague: Kluwer Law International.
13. Sharari, S. A. (2006). Intellectual Property Rights Legislation and Computer Software Piracy in Jordan. *Journal of the Social Sciences*, *2*(1), 7–13. doi:10.3844/jssp.2006.7.13
14. United States Patent and Trademark Office. (n.d.). *Trademark, Copyright or Patent*. Washington, DC: USPTO. Retrieved from http://www.uspto.gov/web/offices/tac/doc/basic/trade_defin.htm

Unit .6

DIGITAL LIBRARIES AND E-GOVERNANCE

Complied by: Muhammad Jawwad

Reviewed by: Dr. Pervaiz Ahmed

CONTENTS

INTRODUCTION

The unit is developed to discuss issues related to empowerment of the masses by digitizing information. The efforts by the Government in this regard by establishing information kiosks and knowledge village centres have also been discussed. The role of e-government and e-governance improving the life of public has also been discussed in the unit. The unit also explain the objectives of e-governance and role of libraries in promotion of e-governance.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is e-governance and, how e-governments work?
2. What are the main objectives and role of e-governance in societies?
3. Why digitization is necessary in libraries and, how libraries are promoting e-governance?

6.1 Introduction

Any country that possesses and masters the information it has, easily promotes economic growth through e-commerce, e-government, social and cultural integration and e-health through the appropriate organization and guaranteed free access to information for its citizens. Trained librarians and archivists possess unique skills in selecting, acquiring, processing and administering information. This makes it possible for members of the public to gain access to knowledge and to use it. For many libraries and archives all over the world, the new Information and Communication Technologies (ICTs) are already having a great impact on the acquisition, storage and access of information resources.

Cultural heritage institutions such as libraries, museums and archives require a major physical and technical preservation effort. They must adapt and change with the pace of society and store growing quantities of educational and cultural materials on networks or in some other digital form, and facilitate access to them. This involves converting large quantities of documents from one format to another, often transferring them to some other storage medium. In modern societies, libraries are particularly important as a means of ensuring that all citizens have access to the knowledge and culture they wish. It is extremely important to organize access to their material. Without such services, society cannot be democratic, open and transparent, because it cannot be assumed that all citizens will acquire a wide range of materials.

6.2 What is E-governance and E-government?

E-governance is an emerging trend to re-invest the ways the government work and a new model of governance would be based upon the transactions in virtual space, digital economy and dealing with knowledge oriented societies. In simple words it is the application of ICTs to governmental functioning to accomplish simple, accountable, speedy, responsive and transparent administration in government. But, E-governance doesn't mean mere computerization of all government office operations or government web-sites on the internet. With the new tools, a networked society's government must completely rethink and re-engineer itself. It is complete transformation of the existing style. Although conventionally the prefix 'e' suggests that an activity is electronic but 'e' also denotes efficiency, effectiveness, empowerment, economic- social development and enhanced services.

E-governance is the public sector's use of information and communication technologies with the aim of improving service delivery, encouraging citizens in the decision making process and making government accountable, transparent and effective.

"E-Government" refers to the use by government agencies of information technologies (such as Wide Area Networks, the Internet, and mobile computing) that have the ability to transform relations with citizens, businesses, and other arms of government. These technologies can serve a variety of different ends: better delivery of government services to citizens; improved interactions with business and industry; citizen empowerment through access to information; or more efficient government management.

E-Government is part of the government's administrative agenda aimed at making it simpler for citizens to receive high quality services while reducing the cost of delivering those services. It applies information technology in the provision of government information and services with an aim of minimizing the burden of public administration and the business activities to its citizens. It also plays a significant role in bridging the gap between the urban and rural population in any country. With e-government, information about the services offered by the government through the different ministries can be disseminated fast as far as the grass-root level through public funded digital libraries and archives. This is intended to avail and improve the delivery of government services to the citizens wherever and whenever needed.

Having e-Government at the centre of disseminating information at all levels (global, regional, national & local) is a way of tapping unrealized potential for high quality government. It emphasizes that while the centre of government needs to create the right conditions for e-government, it is the role of government agencies to actually deliver government information and services through the use of information technology.

6.3 Objectives of E-Governance

The objectives of e-governance are to enhance efficiency, provide total transparency, high responsiveness, accountability, convenience and accessibility of services and information from the point of view of citizens. From governments perspective it is to abolish corruption, cut in costs, an antidote to 'Parkinson's law'

i.e., "works expands to fulfill available". Instant quick reactions in time crisis, easy access to data bases for officials as well as keeping an eye on subordinates and availability of services to public anytime, anywhere. In fact, the purpose of e-governance is to develop "an IT driven system of governance that works better, costs less and is capable of serving the citizens as never before". It is to promote good governance and democracy by ensuring participation of and feedback from the public and making the administration SMART (simple, moral, accountable, responsive and transparent). The objectives of e-government in most cases are set to achieve efficiency in delivering government information and services to citizens; to promote productivity among public servants; to encourage participation of citizens in government; and to empower citizens in line with development priorities.

The major objectives of e-government can be summarized as being to

- improve collaboration between government agencies by minimizing duplication of services and enhancing efficient and effective utilization of resources;
- improve competitiveness by providing timely information and delivery of government services;
- reduce transaction costs for the government, citizens, and the private sector through the provision of products and services electronically;
- provide a forum for citizen participation in government activities;
- ensure that services are available to citizens over a wide range of access tools commonly used by citizens and businesses.

From the above objectives it is clear that e-government aims to utilize ICT to effectively deliver public information, government services and transactions to the convenience of citizens, business and the global community. Apart from servicing citizens, e-government helps in enhancement of transparency, accountability, and good governance, and is attainable through equipping government personnel at all levels with relevant training.

6.4 Role of E-Governance

Network technology has made virtual reality possible, wherein a person can work sitting at home; moreover, in all information processing systems, communication and interaction are facilitated. The application of ICT has been found to be highly

useful in governance, which is known as e-governance, or electronic governance. It plays the following important role:

- Increase efficiency by automation, computerization and networking.
- Supports effective decentralization decision-making by providing an efficient information flow.
- Increase accountability of public services to agencies to citizens.
- Improves resource management.
- Offers the various departments and agencies involved in public service provision, the facility of effective interface with the citizens.
- Increase the accessibility of individual citizens to information and services and allow them to influence government operations.
- Provides comprehensive database which helps policy makers to design, formulate and evaluate policies.
- Facilitates the strategic planning processes which help organization to clearly lay down the objectives, goals, programmers' and projects.
- Enables reduction of paperwork with the use of e-mail and electronic data interchange.
- Enables marketization by supplying information related to the market and enhance public service.

6.5 Why Digitizes Libraries?

Digitization has proved possible for nearly every format and medium presently held by libraries and archive centres. Managers of cultural institutions and those responsible for policy matters related to digitization are often faced with the challenges of understanding the new technologies and their implications for their collections, their institution, their patrons, and the public.

According to the Wikipedia, the term "digital library" is diffuse enough to be applied to a wide range of collections and organizations. To be considered a digital library, however, an online collection of information must be managed by and made accessible to a community of users. Wikipedia further states that digital

preservation is the management of digital information over time. Digital information has transformed the way we learn, the way we communicate, even the way we think. It is also changing the way that libraries and archives not only work, but, more fundamentally, the very work that they do. Information and communication technology, the Internet and digitization have an enormous impact on society in promoting social equality of knowledge. People are able to access the same information, making it easier to define knowledge and culture as common values.

6.5.1 Advantages of Digital Libraries

The advantages of digital libraries and archives as a means of easily and rapidly accessing books, archival materials and images of various types are now widely recognized by commercial interests and public bodies alike. These advantages can be outlined as follows:

- Traditional libraries and archives are limited by storage space. However, through digitization much more information can be stored, simply because digital information requires very little physical space to contain it. This greatly reduces the cost of maintenance.
- A traditional library must spend large sums of money paying for staff, book maintenance, and additional books. Digital libraries do away with these costs. Digital libraries may be more willing to adopt innovations in technology, providing users with improvements in electronic and audiobook technology as well as presenting new forms of communication such as wikis and blogs.
- The barrier of physical boundary is overcome. People from all over the world can gain access to the same information, as long as an Internet connection is available.
- Limited access time is replaced by round the clock availability. A major advantage of digital libraries and archives is that people can gain access to the information at any time, night or day.
- The same resources can be used simultaneously by a number of institutions and patrons. This may not be the case for copyrighted material: a library may have a licence for "lending out" only one copy at a time.
- Information retrieval is enhanced as the user is able to use any search term (word, phrase, title, name, subject) to search the entire collection.
- Though digitization is not a long-term preservation solution for physical collections, it does succeed in providing access copies for materials that would otherwise fall to degradation from repeated use.

- Certain characteristics of objects, primarily the quality of images, may be improved. Digitization can enhance legibility and remove visible flaws such as stains and discoloration.

6.6 ICT and its Relationship to Governments

ICT refers to diverse set of technological tools and resources used to communicate, create, disseminate, store, and manage information. ICT is an umbrella term that includes any communication device or application, encompassing: radio, television, cellular phones, computer and network hardware and software and satellite systems, among others. In government offices, these technologies are connected to the Internet, with the aim of fulfilling information processing and communications functions such as videoconferencing, e-commerce, e-health and e-learning. Over the years ICT has emerged as an important medium for computing, communication and exchange as well as a tool for development by governments at all levels. These are facilities and features that variously support a government's range of activities. They support most government operations to citizens through engagement of its citizens in provision of government services. According to Sato (2003) almost any country or region has a national ICT plan, programme or strategy. However, bridging of the digital divide should be a priority for any country. In most developing countries there is a gap between the urban and rural population be overcome. Thus, each country/region must create its own "bridging a digital divide" programme. Effective use of ICT tools to support e-government services heavily depends on the availability of affordable, reliable, relevant technology infrastructures, skilled manpower and a well-developed national ICT policy.

6.7 Role of Libraries in Promotion of E-Governance

Public libraries are called public Universities as they play a vital role in informing and educating the public during and after their formal education. Libraries in a free society always perform the fundamental function of keeping the public well informed, which is a pre-condition of e-governance. For quite a long time government has used libraries to inform people of their work and policies. Libraries are essential to the free flow of ideas and to maintaining, increasing and spreading knowledge.

Libraries can play an important role in the promotion of e-governance as follows:

- Already established libraries especially public libraries can be used as information centers or public kiosks in the absence of basic infrastructure. Internet connections and other equipment can be provided to them on subsidy basis and the public should be allowed to use them by becoming members of these libraries as is done in several advanced countries like U.S.A., Canada. etc.
- These libraries can work as depository libraries for all types of government information from where the public can access and monitor the work of its elected officials and policy makers.
- Libraries can also help to bridge divide in the digital contents by providing special services to the general public.
- Libraries ensure to freedom to speech, the freedom to read and freedom to view. When the people are better informed, they are more likely to participate in the political process, thus keeping an eye on their elected representatives and making them more accountable.
- Local and state government needs access to information from comparable jurisdiction in order to do their work effectively and efficiently.
- In the field of information literacy, library professionals can play a pivotal role. Besides acting as facilitators and tutors, they can also act as aspirators' and help along with teaching community to spread the culture of information literacy.

The main trust in this area should go to the rural public library. A village or a village cluster with an adequate population should have a community library which will also serve as an information center. Resources of different agencies engaged in the work of public health, adult education, self government, and such others may be pooled to build up this composite center.

According to the Library Association, Britain, "Community Information Services assist individuals and groups with daily problem solving and with participation in the democratic process. The services concentrate on the needs of those who do not have ready access to other sources of assistance and on the most important problems that people have to face, problems to do with their hoes, their jobs and their rights."

There are two major roles that public libraries play in the delivery of e-governance.

- The first, information access, involves working in partnership with other areas of the organization or other tiers of government to deliver and distribute information personally, seamlessly, wherever and whenever it is desired.
- The second bridging the digital divide, is achieved by providing access to information, developing training programmes in the community, ensuring that all people have to access to information, and by understanding the impact of new technologies in the life of people.

6.8 Community information services via the E-Governance

Services provided at the villager's door steps via the e-governance route can presumably cut out the inherent social biases. India really is a land of contrast. We still have a huge section of the society living in the dire poverty, utter ignorance and illiteracy. The important task of e-governance is to remain in touch with the concerned authorities in villages, blocks, tehsil and districts and even in the Metropolis. ICT has made services amenable to people living in cities and villages. The natural progression of information technology—particularly advances that have occurred during the past two decades—along with federal- and state-level e-government initiatives and laws have directly transformed the way governments, library organizations, and other public and private institutions collect, manage, preserve, and disseminate government information to the communities they serve. Government information has evolved from the traditional print format to a growing technology-rich digital-format and online-delivery information system.

This system includes resources such as traditional print and digitized documents, website content, communication technology content ðsocial media and e-mail, and online services e.g., individuals apply online for and manage assisted living services, social security services, and student loans, and they pay taxes and local bills. This technology transformation also creates the need for public libraries to revaluate how they develop, manage, and deliver government information services, resources, and programming to the communities and organizations they serve, address staff training and other library issues related to the delivery of the

information, and redirect scarce resources from traditional services and programming to meet their patrons' government information needs.

Democratic self-governance requires active public interaction and engagement with governments. E-government is a statutory-driven government practice that has the potential to effectively and efficiently increase democratic self-governance by providing technology-enhanced opportunities for public interaction and engagement. Basically, e-government is an accepted government practice that occurs at federal, state, local, and international levels; it has the potential to effectively and efficiently increase provision of and public access to government information through the use of technology; and it has the potential to increase democratic self governance through public interaction and engagement with governments.

6.9 Conclusion

E-governance is step towards reform in government and administration. These reforms focus on bringing improvements in the service delivery, dissemination of information, transparency, public and private partnership, efficiency and accountability. There is an urgent need for sustainable development, In developing countries like Pakistan and India, it is not easy to fulfill all the pre-conditions. But it does not mean we should not go ahead with e-governance. For the successful working of democracy, the empowerment of its citizen is a pre-condition, which can only be achieved by making people fully informed and aware. E-Governance is also seen as a multi- dimensional concept an IT driven methodology that improves efficiency in administration, brings about transparency and leads to reduction of costs in running the government. There are three main domains of e-governance: improving governmental processes (e-administration); connecting citizens (e-citizens and e-services); and building external interaction (e-society). The natural evolution of technology, the Internet, and e-government initiatives and laws have transformed how governments provide public access to information, services, and resources.

Federal and many state e-government laws require agencies to produce information in digital formats and to provide electronic access to the information. In addition, many state and local governments that are not bound by e-government laws recognize the benefits of providing access to their services and resources online. The transformation to an online environment for government information, however, has directly and negatively affected the capacity of public libraries to meet the government information needs of their communities. Each public library community includes patrons who lack adequate access to the Internet for their government information needs, and library tradition includes strategies to provide access for those who need it. Public libraries plan and build their technology infrastructures based on available resources physical space, staff skills and time, funding sources, and system maintenance and patron needs.

E-government changes to the government information environment, however, have altered public libraries' access strategies and tax library networks and available resources. Agencies have moved their services and resources to their websites and closed local offices, which has led to a funnelling effect that sends individuals who would normally visit an agency office to meet their government information needs instead to libraries. The influx of patrons from agencies to libraries has overloaded the capacity of many libraries to provide sufficient Internet access to meet all user needs, and studies have shown that many of these libraries are approaching Internet access and computing service capacity.

Self-assessment Questions

1. Define e-governance and discuss its role and objectives.
2. What is digitization of libraries necessary? Writes advantages and role in the e-government.
3. Discuss the role of ICT and its relationship to governments.
4. How libraries can play an important role in the promotion of e-governance?

Activity:

1. Visit any public library and discuss with the chief librarian, how they are promoting the e-government initiatives? Write in detail the answers in your note book and discuss these answers with your tutor.

Recommended reading:

1. Abdul Kalam, A. P. J. (2004). Inaugural Speech, "Digital Library and its Multidimensions." In *International Conference on Digital Libraries (ICDL-2004),* 24-27 February 2004, New Delhi.

2. Abdul Kalam, A.P.J. (2005). *Address to the Nation by the President of India.* Delivered on the eve of 56th India's Republic Day – 2005.

3. Acharya, S. S. (2004). Sustainable Agriculture and Rural Livelihoods. In Sundaram, K. V., Moni, M., & Jha, M. M. (Eds.), *Natural Resources management and Livelihood security: Survival Strategies and Sustainable Policies*. Delhi: Concept Publishing Company.

4. Ahmad, R. (2002, December). Forging a Sustainable Development Model: The Malaysian way. *Development*, *45*(4), 74–79. doi:10.1057/palgrave.development.1110409

5. Borgman, C. L. (1999). What are digital Libraries? Competing Versions. *Information Processing & Management*, *35*(3), 227–243.

6. Moni, M. (2002). *Digital Libraries in Rural India: A digital opportunity for sustainable development.* Invited talk on the National Technology Day, May 11, 2002.

7. Moni, M. (2002). *Digital Opportunities for Responsive Administration in India: Electronic Administration of Services to Citizens.* Internal Note sent to e-Governance Division of Department of Information Technology, 2002.

8. Moni, M., & Vijayaditya, N. (1990). *DISNIC – ANICNET Based District Government Informatics Programme in India*. Hyderabad, India: Indian Computing Congress.

9. Muddiman, D. (2001). Open to all? The Public Library and Social exclusion: Executive Summary. *New Library World*, *102*(1163/1164), 154–157. doi:10.1108/03074800110390626

10. *World Bank. (2003). Multi-Stakeholder Regional Consultation for International Assessment on Role of Agricultural Science and Technology in reducing hunger, improving rural livelihoods and stimulating environmentally sustainable economic growth. Organized by The World Bank, 12-13 May 2003, New Delhi.*

Unit. 7

BRIDGING THE DIGITAL DIVIDE

Complied by: Muhammad Jawwad

Reviewed by: Dr. Amjid Khan

CONTENTS

INTRODUCTION

The unit is designed to discuss a serious issue "Digital Divide" which is affecting the society due to unequal access to information. Different issues related to digital divide as well as overcoming the same have been discussed in different contexts in this unit. Examples of libraries putting efforts to overcome digital divide have been given. The role of internet and ICT as a communication technology has also been discussed. The government role in respect to diminishing the digital divide in society through promoting self-learning and information literacy programs are also discussed.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is digital divide?
2. What role the Internet and ICT is playing in the digital divide?
3. How can government diminish the digital divide?
4. What is the relationship between human development and digital divide?
5. What is the role of libraries and librarians in diminishing the digital divide?

7.1 Introduction

Although in recent years the information environment has greatly improved in many developed and developing countries, there is no indication that the digital divide is decreasing. Access to information and communication technologies is the crucial issue to a sustainable agenda of socio-economic development. Access to new technologies will furnish vital knowledge inputs into the productive measures of developing countries, especially those who are rural and poor. For the last ten years the term 'digital divide' has become a familiar way of expressing the wide variations in access to information and communication technologies (ICTs) across the world. While there may be some disagreement over what the expression actually means, and what indicators should be used to map it, there is no doubt that developing countries, lag far behind industrialized countries in their take-up of new digital technologies, especially the Internet.

The emphasis in early accounts of the digital divide was on inequalities in infrastructural provision and access to technology. Many countries have devised information strategies, taking for granted the relationship between ICT diffusion and economic development, in an attempt to close the gap. These have generally focused on infrastructural improvements: policies in education, for example, have concentrated on hardware provision with much less attention paid to the issues of how new technologies are used. There is no doubt about the fact that there are many benefits associated with ICTs, but still significant barriers to its effective use exists in both developed and developing countries. Some of these barriers may be endemic (e.g. the generation gap, learning processes and gaining in ICTs, poor telecoms infrastructure, poor computer and general literacy, and so on). Information and communication technologies (ICTs) can be considered as an important weapon in the war against world challenges. When used effectively, it offers huge potential to empower people in developing countries and disadvantaged communities to overcome obstacles, address the most important social problems they face, strengthen communities, democratic institutions, a free press, and local economies and maybe above all, facilitate information flow with which real information society can come true. But, a digital divide separates those who can access and use ICT to gain these benefits, and those who either do not have access to such technology or who are unable to use it for one reason or another. The digital divide

has become a common metaphor originating from now nearly obsolete phrases such as “information haves and have not’s” and “information rich and information poor”. There are numerous approaches that libraries can and have taken in diminishing this divide.

Libraries can play a vital role in bridging the digital divide by providing access to computer and the internet to those who do not have such facilities, but it is not a job that can be done singlehandedly. It is a challenge that must be dealt with in partnership among the various public and private sectors of society, such as the telecommunications industry, educational systems, policy makers, and community organizations. As a result, the present article focuses on some of several dimensions of the digital divide and discusses about some key factors which have an impressive role in digital divide.

7.2 Definition of Digital Divide

According to World Information Society Report, 2007Digital divide is a familiar and dynamic concept which has evolve over time. There is not a single divide, but multiple divides: for instance, within countries, between men and women, between the young and the elderly, different regions etc.

The Internet has profoundly changed many things in our world. The “have” versus “have not” distinction traditionally refers to “the great divide” in our world, between people who have health, wealth and opportunity versus those who have not. The same distinction can be made about people who have access to the new information and communication technologies (ICTs) versus those who have not, and this is referred to as “the digital divide” and globally, there is a gaping digital divide between the “information rich” and “information poor**”.**

The digital divide has been frequently defined. Here, some definitions relevant to our discussion are considered as follows:

- According to Scrutiny of Acts and Regulations Committee, 2005 the lack of access to information and communication technologies to segments of the community. The digital divide is a generic term used to describe this lack of access due to linguistic, economic, educational, social and

geographic reasons. The gap between those people and communities who can access and make effective use of information technology and those who can not the growing gap between those parts of the world which have easy access to knowledge, information, ideas and works of information through technology and those who do not.

- According to Mutula, 2007 a term used to describe the discrepancy between people who have access to and the resources to use new information and communication tools, such as the Internet, and people who do not have the resources and access to the technology.

- According to Alextbox, 2006 the "digital divide" refers to the fact that certain parts of the population have substantially better opportunities to benefit from the new economy than other parts of the population. Most commentators view this in purely economic terms. However, two other types of divides will have much greater impact in the years to come.

7.3 The Role of Internet and the Digital Divide

The internet is a communication technology with a widespread use all over the world. As with other ICTs, the internet provides a potentially powerful means of accessing information. In developing countries the internet offers a cheap and versatile mechanism of connecting users with a global repository of information.

Application of internet technology has transcended all fields of human development in recent years. It has had a wider application in the educational development of all nations, of both the developed and developing world. The new communication technology is said to provide access to global knowledge resources, since it allows users to share their knowledge, resources and experiences of different people. The internet, through the World Wide Web, offers efficient access to research information anywhere, anytime and in any form; which is of paramount use in the conduct of scientific and ttechnological research in developing countries. The Internet acts like an integrator (due to its transcended national boundaries and influences cross border flow of education, health and trade services) and divider (due to its disparities in access).

The digital divide becomes more important because the Internet is not only a communication medium, but also a market place. The Internet gives value to the

customer, enhances profit for producer and eliminates middlemen. The Internet continues has to become more popular and at present has turned into an economic activity. The five areas that need to be given priority for bridging digital divide and converting into digital dividends are distance education, telemedicine, job matching, local development and market reach.

Internet and other computer networks emerged rather unexpectedly. They took on a life and a complexity of social action that governments and most other power structures never predicted their use, for many years, was limited to such a small portion of the population that their social significance was generally not realized until long after the fact of their establishment and unchecked growth.

According to a Foundation survey 2002 the Internet users account for only 6% of world's population and out of that 85% of them are in developed countries where 90% of all Internet hosts are located. This is the essence of global digital divide that needs to be transformed into global digital opportunity. For the Internet to be a true mass medium, it will have to achieve harmony among all consumer segments.

There are different dimensions to digital divide such as economic level of individuals, economic prosperity of nations, ethnicity, age (young/old), rural/urban, gender, geographic location, quantitative and qualitative aspects, dial-up and broadband access. The facts about digital divide based on global perspective reveals an estimated 429 million people are online globally (represents 6% of world population) and out of that, percentage wise, 41 in North America (USA and Canada), 27 in Europe, the Middle East and Africa and 20 in Asia Pacific. Even among highly developed nations, there exists a vast difference in the availability of home Internet access. Sweden ranks the highest (61%) home Internet connections where as Spain has lowest (20%) homes connected. Also, 57% of those not online in the USA have no intention of going online; worldwide, this figure accounts to 33%. Potentially the effect of the Internet in broadening and enhancing access to information and communication may be greatest in poorer nations.

The Internet also offers promise in the delivery of basic services like education and health information to far-flung regions. Distance learning can widen access to training and education. The Internet promises to level the playing field and strengthen the voice of the voiceless in the developing world. In the last few years, international agencies like the World Bank, United Nations Development Programs (UNDP) and etc., have expressed growing concern that the explosion of the Internet may leave many nations far behind, producing growing disparities between advanced industrialized and developing societies. As a result, poorer societies can become increasingly marginalized at the periphery of communication networks.

According to Information have`s and have not`s, 2007 one of the key challenges facing development practitioners today is how to utilize information and communication technology (ICT) tools to provide service to this underserved segment with the objective of decreasing human poverty and improving livelihoods.

7.4 The Role of ICT and the Digital Divide

In the global economy, information and communication technologies (ICTs) play a vital role: International trade is highly dependent on the accumulation and dissemination of information and knowledge; In many countries, the business environment can be vastly improved as business licensing and other bureaucratic procedures are streamlined thanks to e-governance; Information infrastructure thus weighs heavily in the decision of international companies as to where to set up new business and investment; ICT enables developing countries to become active participants in international trade and integrate into the global knowledge-based economy; In particular, mobile technology has enabled farmers and fishermen from rural communities to gain price information and new customers without the need to go through the middleman; ICT also allows many developing countries to make a leap towards higher value added production and services. For example, India and China have become strong players in the ICT industry in terms of both software development and manufacturing of ICT goods such as mobile phones and computers.
In terms of ICT skills, the literature increasingly promotes the need to go beyond teaching IT skills, and emphasizes e-literacy or media skills.

These promote developing peoples' ability to analyze, evaluate and create information. A shift towards a more learner focused approach and teaching a wide range of e-literacy skills is also encapsulated in two recent documents – the ICT Skill for Life standard and the ICT Skill for Life Core Curriculum. However, it appears that the link between funding and accredited courses results in the necessity to follow specific curricula, with little time for learner-led activities or discussions.

Society today is characterized and dominated by applications of modern information and communication technology (ICT) virtually in all aspects of human

endeavors for information access, processing and dissemination. Consequently, the impact of ICT utilization has cut across economics, politics, education, medicine, and other fields of human development. Through the use of ICT, modern society has been described as an information age and the knowledge economy. Thus access to ICT, the internet technology in particular, has provided people with a foundation for building up and applying knowledge globally and particularly in developing nations.

The World Bank (2002) in its report contended that, in the rapidly evolving electronic environment, developing countries face opportunity costs if they delay greater access to and use of information infrastructure and information technology (IT), which together make up ICT. According to the report, ICT is the key for economic development and growth; it offers opportunities for global integration while retaining the identity of traditional societies. ICT can also increase the economic and social being of poor people, empower individuals and communities; and enhance the effectiveness and efficiency of the public sector. This refers to the availability as well as affordability of computers with multimedia performance capabilities and with broadband connectivity to the internet. In a time where so much content and contact is on the internet, anything less than broadband access to multimedia can be considered a competitive disadvantage.

UNESCO Report 2005 rreaffirm the strong correlation (although they concede this does not necessarily mean causation) between the ability to communicate over distances and time using technology and economic growth. The digital divide, while pass in its prognosis, remains a demarcation between ‘‘haves’’ and ‘‘have lesses’’. Among others, caution against the fallacies of present policy measures that tend to favor the economic or intellectual elite in the mistaken belief that this would be more impactful. Therefore the concept of ICT accessibility has to be broadened to ensure that there is greater deliberative democracy, which is essentially what is required in a knowledge society. In this sense, universal suffrage for ICTs means that the availability and cost of access to broadband internet allows a knowledge society to presume that participation is open to all who need them.

The importance of information and communication technologies (ICTs) as powerful tools for socio-economic development is now widely acknowledged not only among large corporations but small business enterprises as well. However, for

ICT to be effectively deployed as engines of economic ddevelopment existing IT skills gap both in developed and developing countries must be addressed. Additionally, pervasive use of ICT in the economy depends on well-trained human resources for developing relevant applications, supporting and maintaining systems. Moreover, investment in human capital, research and development is becoming increasingly recognized as a critical factor in preparing citizens to participate in the digital age.

The digital divide is information inequality which can be reduced through information accessibility, information use and information sharing as well as dissemination. Given that ICT infrastructure is partly prepared, information society citizens can rely on librarians' roles in the increased utilization of information. Furthermore, librarians can help people interact intimately as well as efficiently with new information technologies to locate, use, process, organize, create, communicate and manipulate information and information resources and somehow find an appropriate identity and status in the eelectronic virtual world.

7.5 Information Infrastructures and Digital Divide

Many countries in the developing world still lack the most basic forms of information and communications infrastructure. There are enormous gaps between the technologically advanced, industrialized societies of the world and the developing nations in the availability of information and communication services. Of all the gaps that exist between the South and the North, none is growing faster than the information gap.

According to Information have`s and have not`s, 2007 infrastructure is one of the key foundations for building an information society and bringing the benefits of ICTs to all. Especially, public intervention in facilitating the establishment of the basic network infrastructure is vital. Such public infrastructure would reduce cost of entry for new companies into the market and substantially reduce the risk of business. Both of these would also reduce cost of information for the general public.

Basic infrastructure includes buildings and power resources; many countries lack power generation and distribution facilities adequate to running computers or Network Infrastructures except in large cities, and access can be limited and/or

sporadic even there. Solutions involving the use of small scale local power generation (solar power and fuel cells) are becoming increasingly workable, but a high bandwidth Network Infrastructure often depend on the backbone provided by a power infrastructure.

The Internet is built on a complex layering of data networks, with a variety of top level Network Service Providers (NSPs) interconnecting a wide variety of localized networks, including schools, business, governments, and local Internet Service Providers (ISPs). Individual local ISPs will provide end users and smaller businesses with one or more of a variety of connectivity options, including dial-up modem access, broadband cable network access, radio frequency network access, and others. End users and small businesses will, in many cases, use this connectivity to connect multiple machines to the Internet via a local area network (LAN). Although, in future scenario for internet usage, developed countries will continue to offer a solid consumer base for ICT products and services, the future expansion of the ICT industry will be driven by growth momentum in developing countries. Thanks to their younger age structure, larger populations and lower labour costs, manufacturers of ICT goods (e.g. mobile phones and computers) will be increasingly attracted to developing countries to invest and set up production. An increasing number of developing countries will improve their business environments through e-government and e-business. This will help to facilitate national and international business transactions and boost economic growth in these counties.

7.6 The role of Government and E-government in Diminishing the Digital Divide

According to E-government Report 2007. E-government is a common term used for the concepts like e-services, e-democracy and above all, e-administration. It is the use of information and communication technology to increase the access to, and delivery of the government services to benefit citizens, business partners, public and private sectors. E-Government engages the automation or computerization of existing paper based procedures that will rapid new methods of management. It is a connection between the government authorities, the private sector and the citizens. Above all, it aims to advance the government's resources in the direction of an effective control to increase the country's social, political and economic resources for development.

E-governance projects, can bridge the digital divide and spread the benefits of ICT especially to rural citizens, enable them access to digital services and help address and hopefully eliminate the rural/urban digital divide. (E-Government projects will

bridge digital divide, spread ICT benefits) Connecting people by making investments in the technology is comparatively an easier job than bringing and keeping them online. The key to encourage citizens in developing countries to use the Internet would be to provide them with compelling content and services that meet their primary needs. E-government perse is an unlikely key for bridging the digital divide. The notion that technologies can prescribe their own course of action is mythical. The responsibility for technological outcomes lies in the social order between individuals, groups and institutions through which lives are organized. There is, however, synergy to be created between technology and social context.
This is not deterministic synergy it will vary by context and will therefore entail adaptation of the technology and social context. If there is no prior experience of using computers, then there aspects, this may make it easier to introduce new technology, as there will be no old values or context to change; however, a context has to be created. If the existing context for information is informal or non-paper driven, automation may be difficult. In both situations, new rules will have to be learnt and accepted. When leapfrogging, one must be careful to identify both technical and social considerations, ensuring that the technology is not embedded to the detriment of the social order.

Further, one should not assume that there is only one way to leapfrog all social contexts will have one or more different leapfrogging solutions. By most measures, developing countries continue to struggle with the implementation of viable e-governance strategies.

Information can be used in various ways some being more productive than others. Hence the demand for information (the willingness of individual
to pay a price for it) also depends on the ability of the individual consumer to utilize information. It also depends on the consumer's awareness level of the availability and source of information. It can be argued that government has a role in enhancing awareness and capability of the people to use information properly and thereby increase demand. Such enhanced demand would create a business case for the market to exist.

In addition, E-Government is about a process of reform in the way governments work, share information and deliver services to external and internal clients for the benefit of both government and the citizens and businesses that they serve. E-

Government harnesses information technologies, such as Wide Area Networks (WAN), Internet, World Wide Web, and mobile computing by government agencies to reach out to citizens, business, and other aims of the government to:

- Improve delivery of services to citizens
- Improve interface with business and industry
- Empower citizens through access to knowledge and information, and
- Make the working of the government more efficient and effective.

The resulting benefits could be more transparency, greater convenience, less corruption, revenue growth, and cost reduction. So in order to bridging the digital divide, NGOs and grass root organizations that catalyze and manage the community building process needs to be improved. Applications that draw a large cliental that pays for the information services, ensure economic viability that empowers rural citizens and enables formation of communities Technology that makes rural access inexpensive and robust.

According to Asia-Pacific Human Development Report, 2008 Many different institutions, some related to incipient state organizations and others to private associations, emerged to provide the public goods required for the proper functioning of markets and for reducing transaction costs. (The knowledge explosion and the knowledge divide) It is a fair generalization that thus far the generation of technology has been confined to a few countries and is becoming increasingly capital intensive, making it more difficult for developing countries to catch up. Also more and more R&D is increasingly originating in the private sector. The private companies that produce new goods and services that improve human outcomes are motivated. These companies expend huge amounts on R&D and capture the returns to such expenditures by staying ahead of the competition. These private institutes ease the communication between government and people and help people getting information faster and easier and getting more information. The qualification of IT and ICTs in this process, help diminishing digital divide and the role of it is inevitable. Increasingly, global attitudes toward technology are being shaped by important non-governmental organizations (NGOs), especially as non-state actors start to play a greater role in global governance. As yet, there is relatively little research on the social impact of privatizing social services. Most investigations have focused on economic efficiency rather than impact on redistribution

and the poor. If policy is weak before privatization, it will also be weak after privatization. Privatization is clearly no substitute for a responsible policy of redistribution.

As a result, we can conclude that in market economies, both the public and private sector can promote digital opportunity. Government has an important role to play in establishing an enabling environment for investment and market competition, as well as intervening to achieve socio-economic goals in areas to create balanced growth in reality. The roles of governments and the private sector overlap and include additional elements: first positive incentives stimulating market dynamics and second, measures are preventing
uneven development (for example by use of regulatory criteria or restrictions).

7.7 The Relationship Between Human Development and Digital Divide

7.7.1 Human Resource Development

The importance of information and communication technologies (ICTs) as powerful tools for socio-economic development is now widely acknowledged
not only among large corporations but small business enterprises as well. However, for ICT to be effectively deployed as engines of economic development existing IT skills gap both in developed and developing countries must be addressed. The Digital Opportunity Task Force emphasizes human resources development through systematic training and education as critical if countries have to reap digital divides.

Additionally, pervasive use of ICT in the economy depends on well- trained human resources for developing relevant applications, supporting and maintaining systems. Moreover, investment in human capital, research and development is becoming increasingly recognized as a critical factor in preparing citizens to participate in the digital age. Enhancing the information literacy in developing countries is one of the most important strategies in order to decreasing the digital divide.

7.7.2 The Impact of the Digital Divide on Skills Development

There has been a great deal of discussion about the impact of digital information resources, particularly around what has been termed the digital
divide, or the split between those that have access to digital information resources and those who do not. In general, the digital divide applies internationally. However, (Norris 2001) discusses three types of digital divide: Social (within

countries), global (between countries) and democratic (those unable to use ICTs to take part in public life). The (social) digital divide has an impact upon information professionals in developed environments, with many libraries lacking resources and technical support, and staff needing continuously to acquire appropriate training in order to deliver up-to-date services and troubleshoot equipment.

Regarding the *global* digital divide, Lim, (1999) suggests that too much emphasis has been placed upon the development of ICT infrastructure in developing countries, and not enough consideration has been given to human resource development. However, in order to understand how ICTs impact upon skills development in developing countries, it is necessary to recognize the situation that currently exists regarding the ICT infrastructure. For example, in Africa, one in a hundred people have access to a PC; the few Internet Service Providers are comparatively expensive; power supplies may be unreliable (even non-existent) and telecommunications are sparse, with the 90% of the population living in rural areas having only 50% of the telephone lines . Suggests that many ICT users are self-taught, and are capable of developing an understanding of ICTs through the experience of utilizing them. If this is the case, then countries unable to provide extensive access to ICTs are inevitably marginalized as they are less likely to produce capable self-taught persons. However, these suggestion does indicate a more optimistic scenario for those developing environments progressively providing access to ICTs, as it suggests that staff may be able to gain at least some degree of expertise through self-learning.

7.7.3 The Information Literacy Needs Attention

The education system in information age has slight alteration and just some sort of education systems has become popular solely in developed countries. Developing countries are still suffering from digital gap. They can't keep up with the changes in socio-economic system during the twentieth century. So an up-to-date education system should be consistent with current socio-economic system that has named information society, which means it should be based on concepts from information and consumer services. The need for adequate scientific and technical literacy in a nation's workforce, an important element in the development of science and technology capabilities, has been recognized as nations strive to raise their living standards. This is now perceived as an everybody working knowledge of science, is as necessary as reading and writing for a satisfactory way of life in the modern

world. In an era where economic growth based on the production of primary goods is proving to be increasingly uncertain and difficult to sustain, the call for improved human resources that can competitively turn out ever more sophisticated goods and services rings true in rich and poor nations alike. So, science and technology capabilities of most developing countries are far too limited to deal adequately with the challenges of economic advance, social progress and environmental sustainability.

Thus, they should work hard on their human development programs and invest on their education and literacy. (The knowledge explosion and the knowledge divide, 2001) Judging by these experiences, the most problematic issue is lack of support to training and maintenance. This was emphasized even more frequently than cost or lack of infrastructure, by groups throughout the world. “Training, training, training”, says Mike Jensen, one of the first activists in the Association for Progressive Communications. Health Link encourages a ” mentor/ supportive” approach to training, doing as much on site as possible, and following up with repeat visits. Information has its existence, since the existence of human being. With the advent of ICTs and computers, the delivery of information has become easy, inexpensive and every ones cup of tea. As information is increasingly codified in digital forms, new skills are needed to operate the technology to search for, organize, manage information and use it to solve problems and create new knowledge and cultural products. The greatest challenge for society in 21st century is to keep pace with the knowledge and technological expertise necessary for finding and evaluation information. Information literacy is a survival skill in the information age. Instead of drowing in the abundance of information that floods their lives, information literate people know how to find, evaluate, and use information effectively to solve a particular problem or make a decision.

Libraries, which provide a significant public access point to such information and usually at no cost, must play a key role in preparing people for the demands of today`s information society. Technological literacy, a broad understanding of the human-designed world and our place in it, is an essential quality for all people who live in the increasingly technology-driven 21st century.

What are the features of information Society that education systems must be commensurate with ? The information society is a new kind of society. Specific to

this kind of society is the central position information technology has for production and economy. Information society is seen as successor to industrial society. Closely related concepts are post-industrial society, post-Fordism, post-modern society, knowledge society, information revolution, and information society. Applications of information and communication technologies (ICTs) are making dramatic changes in economic and social development that these tectonic economic and social changes have been characterized by terms such as “knowledge economy” and “learning society”, conveying the notion that knowledge and learning are now at the core of economic productivity and social development.

Nowadays information communication technologies are the nervous system of contemporary society, transmitting and distributing sensory and control information, and interconnecting myriad interdependent units. These technologies consist Electronic Mass Media such as Cable Television, pay Television Services, Interactive Television, Wireless Cable Systems, Streaming Media, Radio broadcasting, Direct Broadcast Satellite, Computers and Consumer Electronic such as Multimedia computers Video Games, The Internet and the World Wide Web, Office Technologies, Internet Commerce, Virtual and Augmented Reality, Home video and Digital Audio and Technology and Satellite technologies such as Local and Long Distance Telephony, Broadband Networks, Residential Gateways and Home Networks, Satellite Communication, Distance Learning, Wireless Telephony, Video conferencing.

Education is essential in these parts of the world. Training and improving already existing skills is the first step in establishing any form of economy. Utilizing labor and skills will allow for countries to evaluate and determine what their strongest assets are in the trade world. Once this is established, and a monetary system is developed, a democracy is also needed. With a democratic leadership the people can make decisions for themselves, and provide for their futures. After such communities are established, then technology can start to integrate into the picture. However, substantial amounts of education are needed to train workers on how such technologies work, and how they can benefit the culture. Many of these small countries must rely on trade to be viable. They must learn to utilize their assets and resources adequately. “Improved training, particularly at secondary and vocational levels, will be necessary for the effective functioning of the economies,” With a strong educational foundation, many of these countries will continue to flourish,

while others will begin a new venture with economic growth and stability. Building an economic base, education and training can lead to more substantial technology integration. Producing energy and dispersing this energy is the next step. After power supplies are in tact, communities can begin to slowly integrate technology.

Education is the key to this integration, and without proper training and understanding, resistance of the Western influence is likely. When societies learn to utilize their own strengths and resources, the feeling of ownership leads to determination. This determination along with assistance, allows for integration to begin. As we continue to bring technology into Third World, developing countries and less developed countries, let us keep in mind the main reason for this integration. It is not to control their markets and trade; it is to bring them to a higher level economically and socially. When we begin to integrate strictly for our own gain, the global network will fail. Many countries do not encourage the technology and free market/free trade ideal. They are opposed or skeptical of the Western influence that might also invade their culture. With a tyrannical government, this idea of integration will never work. People in these situations are not valued, and they are not able to acquire rights that allow them to even consider integration and its possibilities.

Poverty is such a huge problem in many of these countries. Unless a political structure is in place, economic development is nearly impossible, and without the economic backing, the poverty will never diminish. Programs implemented in the past by the US and other countries have not always succeeded. Again, this is a case where we must learn from the trials and mistakes of past integration in order to continue in the process today. Not only will the less developed countries benefit, the developed countries will see a boost in political, economic, and educational aspects of all cultures.

The Internet has become such vital part of our society and cultures that is often difficult to imagine how we ever functioned without it. The idea that “knowledge is power” has become possibly one of the most quoted (and east understood) truisms of the Information Society. There are many ways of looking at knowledge or information, especially when we consider the *volume* of information that becomes readily available through computer networks and databases.

Anthony Smith, In his book "*The Geopolitics of Information"* addresses the context in which we think of this continuous stream of information: "It is possible to view information as a social resource of a special kind rather than as a produced commodity, a resource which enables other resources to function productively since it is the existence of salient information which determines the value and existence of other resources." He adds: "The problems of privacy, access, commercial privilege, public interest, are problems of allocation and priority and value of the kind that every society has had to debate incessantly in history and now has to do so again in this new guise." Smith's suggestion that there is"geopolitics" to information itself is an interesting one. It is premise worth examining in some detail. Originally the subject of geopolitics focused upon the analysis of geographical influences on power relationships in international politics". However, technological developments in the later half of this century have allowed many states (and other social organisms) to overcome limitations placed upon them by geographical location. In the modern context, geopolitics is less and less geography bound and today might be better defined as a study of relationships between space and power as these apply to social organisms.

In order for developing countries to accept the Internet and use it effectively, these countries must agree upon some type of policy to ensure success. "Developing countries must create a framework of political, legal and economic conditions that guarantee equality of opportunities and create incentives for trade and investment,". Most of these countries still have political systems that are imperfect to those in more developed countries. "In the rankings of economic freedom, most developing countries obtain poor scores in critical areas such as legal stability, size of government, regulation and sound monetary policy," These impediments often keep these countries in poverty.

With the Internet and world support, it would be possible for these developing countries to break out of this lifestyle. Some of these countries have not yet accepted globalization and they have not embraced economic liberalization. Perhaps with better campaigning and education of the importance of such ideals, the countries would more graciously accept these concepts and help improve their countries. Most of these countries have political institutions that spend too much money and borrow too much to even consider individual rights for its people .With

our attempts at open trade and cultural integration, it is hoped that technology can help break individuals in these countries free from the oppression they face with such rule.

Lack of capital, inadequate markets, and weak institutions are demonstrably inadequate. The crucial element that has been largely ignored is the cultural: that is to say, values and attitudes that stand in the way of progress. Some cultures, above all those of the West and East Asia, have proven themselves more prone to progress than others. The conclusion that culture matters goes down hard. The implication is that all cultures are equally worthy, and those who argue to the contrary are often labeled ethnocentric, intolerant or even racist. A similar problem is encountered with those economists who believe that culture is irrelevant that people will respond to economic signals in the same way regardless of their culture. But a growing number of academics, journalists and politicians are writing and talking about culture as a crucial factor in societal development, and a new paradigm of human progress is emerging but the vast majority of countries still lags far behind. Of the six billion people who inhabit the world today, fewer than one billion are to be found in the advanced democracies. More than four billion live in what the World Bank classifies as “low-income” or “lower middle income” countries.

As a conclusion, with the *proper education, reliable information, and having right to have information,* these communities can develop a democratic society where people are allowed to view and process information and make logical conclusions on their own.

7.8 The role of Libraries, Librarians and Digital Divide

Nowadays, terms such as “knowledge-based society”, “sustainable development” and so forth are frequently heard. In the world today, information is “a resource for development”, and “the absence of reliable information is an epitome of under development” Special librarians, who have access to information, have an important role in developing countries, in order to delivering proper and reliable information to people. So they can affect on promoting learning and information

literacy programs. Libraries will be called upon to provide authentic and reliable information, evolve strategic alliances, participate in network activities and contribute to the bridging of the digital divide. They will emerge as technology experts, guides, researchers, analysts, knowledge engineering, editors, navigators, gatekeepers, brokers and asset managers. Conclusively, librarians will need to acquire new skills to access net resources, and develop new strategies and services to meet the challenges of the knowledge age.

These days, Libraries have more important role than before in developed and also in developing countries. They are emerging as one of the most important vehicles for bridging the digital divide as they increase the availability of technology to patrons and elevate the technical expertise of staff. Providing libraries with computers and librarians with training on how to use the new information technology is one of the most important solutions in order to bridging the digital divide all around the world. Libraries, with their long history of providing free access to information, are a natural vehicle for leading this effort to expand public access to technology. A key component of assuring the sustainability of these efforts is an intern program, which works to faster relationships between various kinds of libraries, specially library schools and to help with computer installation and training. These activities help ensure that every library is prepared to maintain and sustain a high level of technological access for all patrons.

There is strong evidence that this and other similar efforts are making a dent in the digital divide. The U.S. National Commission on Libraries and Information Science (NCLIS) recently issued the results of its 2000 Internet Connectivity Study, which measured the level of connectivity, public access, training support and technology funding, current and anticipated, for staff and the public. The NCLIS found that in the last two years, Internet connectivity in public libraries has increased from 83 percent of libraries being connected to over 95 percent and in the same time period, libraries have nearly doubled the number of public access workstations. Libraries must train and teach their patrons to handle new electronic information formats as never before. They teach computer skills, internet surfing, information searching and providing various electronic services for the local citizens. In the electronic

service context, Librarian's roles will become more prominent as educators, information managers, information management consultants, custodians of information, information providers and publishers, change agents and custodians of library facilities:

- As educators, librarians can increase awareness among their clients of information networks, their contents and potential use.
- As information managers, librarians need access to information resources of many types, in many disciplines. They have the skills to build navigation tools for networked resources in the same way that develop navigation tools for published information in library catalogues and national bibliographies.
- As information management consultants, librarians can help network users build and maintain personal information systems, which provide access to the subset of networked information sources relevant to each user's work.
- As custodians of information, librarians are facing apparent challenge to their role, as physical resources migrate into electronic form and on-demand electronic delivery becomes more common.
- As information providers, librarians can make available much more widely collections which now can be sued only within a single physical library location.
- As change agents, library staff can lobby managers and governments as appropriate, for network access for themselves and their users.
- As custodians of library facilities, library can provide workstations, network gateways, printers and software that may not be otherwise available to the public.

Librarians should confirm themselves even stronger than before thought their knowledge and skills and they should change their attitude towards the new environment, if they want to have an important role in order to diminish the digital divide. In this respect "librarians must redesign their service menu for their customers through thinking functionally. Under the present conditions where the Internet has become very popular, they should concentrate on the professional information services which can be provided only by libraries and librarians. For that purpose, we need to review the various IT devices and information services."

While the digital divide is expanding, the role and responsibility of librarians and libraries are increasing. So policy makers should consider the vital role that librarians can play in the realization of a knowledge-based society and sustainable development. So it is emphasized that proper access to the information, is the key to alleviate the obstacles to information access, and so diminish the educational and information divide and totally digital divide within each country. Yet we must not overlook that there are basic, essential requirements – economic, social, political, and cultural – that must be taken into consideration to overcome the digital divide and thus gain equally access to information.

In this situation, Librarians and librarians have worked to bridge the divide between the information ”haves” and “have- not’s” for more than 100 years. (Digital divide, (2002) Developing countries try hard in order to diminish digital divide. Libraries throughout North America have used resources provided by the Foundation to support their own individual efforts to increase public access to information. In the Yukon Territory, librarians established a mobile computer lab that travels to remote areas not served by a local library. In St. Louis, technology centres were installed in libraries serving inner-city neighbourhoods. Students from the surrounding neighbourhood use the technology centres to do homework after school. In Michigan, local libraries installed the first public workstations with Internet access. As one of the library’s missions, user education has been developed to help build up the learning skills for users regardless of age, race, language, religion, sex and physical ability. The term “information literacy” is evolved from the library education, for instance, the skill of catalogue search.

7.9 The Role of Digital Libraries in Diminishing the Digital Divide

While the information and communication technologies (ICTs) in general, and the Internet and the world wide web in particular, have made life easier by facilitating easy communication with virtually everyone, and easy access to information located virtually anywhere in the world, they have also widened the gap between the rich and the poor. Digital libraries make use of ICT and the web to provide access to the local and remote digital information sources and services. Therefore, accessibility to the basic ICT and the Internet is a pre-requisite to the development and use of digital libraries. There are many other factors too. The two most important issues are the cost of building and maintaining sustainable digital library systems and

services, and achieving the required information literacy standards so as to exploit the full benefits of digital libraries.

Digital libraries can play a significant role in bridging the gap. Developing countries, especially the least developed countries, that struggle to meet the basic human needs, cannot afford to spend such huge amount of money required for research and development of digital libraries. In addition, there are many other problems that stand in the way of digital library development in the developing countries. Consequently, libraries have long been suffering from financial and other crises such as lack of the appropriate technology, trained manpower, etc. Libraries have also been affected by a number of social problems, the primary ones being the poor literacy rates. While governments are struggling to improve the levels of basic literacy, proper use of library and information services call for another level of literacy – the information literacy that is absolutely necessary for people to become good information users. Due to the lack of suitable technologies and trained manpower, and above all due to the lack of financial resources, most libraries in the developing countries do not even have fully developed and up-to-date OPACs, let alone full-fledged automated library management systems, and digital libraries. Hence compared with the developed world scenario, libraries in the developing countries are already left behind by at least one generation.

Digital divide and lack of resources for digital library research and development may increase the gap far more significantly between library and information services in developed and developing countries. Five specific areas where digital libraries can promote developments in the developing
countries:

- In the dissemination of humanitarian information
- In facilitating disaster relief by providing the appropriate information
- In the preservation and propagation of indigenous culture
- In building collections of locally-produced information, and
- In creating new opportunities to enter the global marketplace.

To this we can add another important point that can be applicable to any digital library: digital libraries can facilitate lifelong learning which is the key to success in this fast changing world. The following is a short list of problems or issues that

stand in the way of digital library research and development in the developing countries:

- Shrinking library budget that forces the library management to struggle to maintain a minimum standard of services leaving no room for new ventures and developments
- Lack of financial support specifically for digital library research and development
- Absence of fully developed and up-to-date OPACs, and little access to online information resources – on-line data bases, ejournals, etc.
- Poor ICTs – computers and networks
- Poor facilities for access to ICT, especially the Internet
- Stringent government and institutional policies on Internet access
- Lack of trained manpower
- Poor information literacy rate that causes lack of appreciation of modern information services and their use.

The list may go on and on. Any experienced library manager from a developing country can surely add quite a few more points to the above list. In short, existing libraries in the developing countries are struggling for their mere existence. Of course there are many reasons for the lack of resources for library development. In countries where citizens still struggle for reliable sources of food, water, medical care and educational opportunities, bridging the digital divide may seem like a lofty goal, and that is why digital library development is way down the list of priorities of governments and institutions. As digital libraries become part of the traditional library infrastructure, it will be vital to deal with a number of issues. A major risk to digital objects is technological obsolescence of the devices to read them. Likewise, a major worry is the funding for the regular refreshing. Digital preservation will be an ongoing operation, requiring a regular future expense. Experts have projected that over the next 20 years, with all the potential new ways to create, receive and transmit information, we will receive 50 times as much as we have had in the past. So digitizing our libraries should be one of the proper solutions in order to bridging the digital divide all over the world.

7.10 Conclusion

Closely linked to a nation or society's scientific and technological capabilities and to the issue of scientific literacy, is its people's ability to access and effectively utilize the rapidly growing flow of information. Despite the ongoing knowledge explosion, however, entire nations and millions of individuals, particularly in the developing world, are ill-equipped to be part of an emerging global "information society" due to factors such as inadequate education, social and political exclusion, and sheer lack of financial resources. The digital divide is shrinking in terms of Internet usage around the world and it has reduced sharply over recent years. The prevalence of the digital divide in developing countries is a serious threat to effective utilization of information for socio-economic development and nation building. Governments and relevant international/ non-governmental organizations should introduce appropriate programs that will entice people to consider careers in science and technology. In addition they should provide internet access and technology education for people irrespective of their career in order to bridge the observed digital divide.

One serious form of digital divide in developing economies as shown by the survey is that of educational/ literacy level of the citizenry. In developed nations, the vast majority of the populace are educated and/literate. In developing countries, however, where there is a high incidence of illiteracy, a lot of people are cut off from having access to vital information especially through computers and the internet, unlike in telephony. This is explicit since an illiterate person can only make use of Internet through intermediation. Most of these illiterate people in developing countries are not even aware of the existence or usefulness of ICT. Hence, governments, relevant international organizations and non-government organizations (NGOs), etc., should embark on enlightenment and/awareness programs regarding ICT, particularly the Internet and its applications in different fields of human endeavor. This will encourage the illiterate/uneducated section of our population to begin to make use of available Internet cyber cafes at least through intermediation for relevant and specialized information on business, farming, fishing etc. Then a long-term measure will be for governments to embark on policies that will improve literacy level in developing countries as obtained in the developed countries. In addition, general computer literacy programs should be

organized by public institutions/ministries in the country to reduce the level of the digital divide due to lack of skills in modern computer technology among their staff. Respective governments/ agencies should therefore make the internet facilities available in our public institutions: schools, tertiary institutions, ministries, etc. for ease of access (e.g. free access or subsidized financial cost) Relevant governments and their agencies should also embark on the provision free or subsidized public internet cyber café in developing countries, as well as rural communities, to bring Internet access to the low income earners.

It also recommends the following five strategies to help make progress toward digital equity:

1. Legitimize the significant role culture plays in students' educational experience.
2. Continue to challenge perceptions about the role of technology in education.
3. Encourage others to recognize the critical link between technology, professional development and classroom practice.
4. Create opportunities for students to access technology outside the classroom.
5. Continue to seek funding for technology in spite of challenges.

So, the Internet and related technologies, generally information communication technologies, can provide information and tools that extend the fastest and newest ways of learning and can establish new kinds of education systems that will foster learning in an interactive, digitalized and hypertext atmosphere and this will be vital task of all the countries over the world to prepare these infrastructures that are useful for both students and teachers. As we know, one of the underlying causes of the digital divide is unequal access to technology and uneven ability and knowledge to implement and use that technology. ICTs are a major force for development in the modern economy, but many people still remain unconnected. Creation of "technological parks" by the central and regional, and municipal governments, creation of "technological business incubators" by the universities and other academic institutions to entrepreneurs interested in exploring new business venues. These incubators would provide technical, business and technological management assistance. Promotion of clusters and networks of medium and small enterprises in specific sectors and localities, specializing in specific aspects of productive

services, which exchange products and services and share a series of supporting services to achieve greater collective and individual efficiency. Promotion of linkage between national firms and transnational firms that would purchase local products to make them available in global markets, promotion of strategic alliances between national enterprises in key sectors for the transformation of national resource, creation and promotion of capacity building and professional training specialized for the productive social priority sectors and etc are some of the strategic and policies to bridge the knowledge divide specially in developing countries.

In addition, there are some solutions but community computer and Internet kiosks have emerged as the preferred medium for bringing the benefits of the information and communication technologies (ICTs) to rural communities in developing countries. Countries need to have access to lower cost computers and they need to have easy and affordable access to the phone lines and Internet. As a more developed country, we need to learn from the successful and unsuccessful attempts at integrating technology. Not only will these less developed countries benefit from the integration, developed countries will benefit politically, economically, and educationally.

The digital divide can never be contained in isolation but the effort has to be multi-dimensional and multi-pronged. ICTs are one of the enabling tools to bridge digital divide. Creation of ICT infrastructure and content are core methodologies and a thrust to technology growth in a planned manner will certainly lessen the gap. While digital divide is an issue of recent concern, technology divide has been as issue for much longer. There are two approaches to enable a wider population to benefit from technology and information revolutions; one is to enhance level of literacy (basic, functional technology and computer education amongst masses) and another is to design appropriate IT tools around the capabilities of users (such as Simputer that employs audio/visual input/output, without need to be literate; low cost telephony and data communication—VOIP and wireless communication.

Self-assessment Questions

1. Define digital divide and discuss the roe of Internet and ICT in the digital divide.
2. Explain the relationship between human development and digital divide.

3. How can libraries and librarians diminish the digital divide in society?
4. Explain the information literacy need in a digital divide.

Activity:

1. Prepare a report with the help of tutor on the digital divide phenomena exists in Pakistani society.

Recommended Readings:

1. Balasubramanian, S., Valarmathi, C., & Narmatha, T. (2002). Digital Divide: Problems and Solutions. In Kaul, H. K., & Baby, M. D. (Eds.), *Library and Information Networking NACLIN 2002* (pp. 369–374). New Delhi: DELNET.

2. Bridges.org. (2002). *Spanning the digital divide*. Retrieved May 24, 2006, from http://www.bridges.org

3. Chowdhury, G. G. (2004). Access to Information in Digital Libraries: Users and Digital Divide. In *International Conference on Digital Libraries: Knowledge Creation, Preservation, Access and Management, Conference Proceedings* (pp. 56- 64). New Delhi: TERI.

4. DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social implications of the Internet. *Annual Review of Sociology*, *27*, 307–336. doi:10.1146/annurev.soc.27.1.307

5. Grace, J., Kenny, C., & Zhen-Wei Qiang, C. (2004). Information and Communication Technologies and Broad-Based Development: A partial Review of the Evidence. *World Bank Working Paper, No. 12*.

6. Kaul, S. (2002). Digital Divide: Are we experiencing the era of technological ‘haves’and ‘have nots? In Kaul, H. K., & Rama Reddy, E. (Eds.), *Library and Information Networking NACLIN 2001* (pp. 345–358). New Delhi: DELNET.

7. Kiiski, S., & Pohjola, M. (2002). Cross-country Diffusion of the Internet. *Information Economics and Policy*, *14*(2), 297–310. doi:10.1016/S0167-6245(01)00071-3

8. Lal, K., & Paul, S. (2004). Digital divide and international orientation: Evidence from Asia Pacific countries. *The ICFAI journal of Applied Economics, 3*(2), 31-41.

9. Mansell, R., & Uta, W. (Eds.). (1998). *Knowledge Societies: Information Technology for Sustainable Development*.Oxford, UK: Oxford University Press.

10. Oyelaran-Oyeyinka, B., & Lal, K. (2005). Internet diffusion in sub-Saharan Africa: A cross-country analysis. *Telecommunications Policy*, *29*, 507–527. doi:10.1016/j.telpol.2005.05.002

11. Paul, J. (2002). Narrowing the digital divide: initiatives undertaken by the Association of South-East Asian Nations (ASEAN). *Electronic Library and Information Systems*, *36*(1), 13–22. doi:10.1108/00330330210426085

12. World Development Indicators (2008). Washington, DC: World Bank.

Unit. 8

DIGITAL LIBRARY DEVELOPMENT: INTRODUCTION

Complied by: Muhammad Jawwad

Reviewed by: Dr. Amjid Ahmed

CONTENTS

INTRODUCTION

The unit focus on the technological aspects of digitisation. Anyone interested in knowing the details of the techniques and equipment's for digitizing would find this unit very helpful. A student learning digitization as well as a librarian interested in digitizing his/her library will find useful tips in this unit. Example of various digital libraries and scientific societies or organizations are also presented in the unit in detail.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is digital library development and organization?
2. What documents digitized, in which formats saved and, how retrieved?
3. How digital libraries and documents collection preserved?
4. What are different types of libraries and their contents?

8.1 Introduction

A library is a place where information is disseminated professionally to needy clients in a user-friendly environment. The application of computers to the various tasks by which information is collected, processed, organized and showcased was a significant development in the latter part of the last century. Computerized processes gave rise to dynamic as well as interactive products such as the inventory/database of the collection that can be searched and accessed electronically, e.g., the online public access catalogue (OPAC), in contrast to the comparatively static card catalogue. The single most important development that has brought about sweeping changes in the library and information discipline in the developed world is that of digital libraries. Though the professionals and libraries in developing countries are also experiencing the virtues of the Internet and electronic information highways, many of these libraries have not gone much farther than the computerization of in-house operations, making use of databases available in electronic media such as CD-ROMs, and Web access to subscribed journals and various free resources. Some of these libraries created websites basically to project the services and strengths of the library and to serve more as advertising or public relations media rather than digital information gateways. Digital library development should be taken up as an additional task to populate the websites with valuable in-house content like research reports, publications of inhouse researchers, and so on. In a country such as Pakistan, so rich in indigenous research and development in disciplines varying from science and technology to social science, humanities and spirituality, there is a tremendous need for hosting full-fledged digital libraries by appropriately tagging this content with affordable information technology. The recent years witnessed the boom of the Internet the world over, leading to the acceptance of the Web as an alternative delivery mode of information products. You never see a renowned publisher without a website now, and most of the international journals offer some provisions to access the abstracts/full text of papers through the Web along with the print subscription. The information centre is undergoing a transition from the paper-dominated manual environment to the shared access-oriented electronic environment.

8.2 Construction and Organisation

8.2.1 Some Technical Questions

The backbone of any Digital Library is of course computer software, hardware and the reliance on high speed networks. Any Digital Library would rely on components such as local networks with connections to the Internet; databases with user friendly interfaces for searching and administrative purposes which can index full text documents for fast access.

The OPACs (Open Public Access Catalogue) of most libraries are systems that embrace all kinds of content and functions of the organisation in one single system. This is not necessarily true for the Digital Library that most likely is a collection of different systems and resources connected through a network and searched and reached by a web-interface

Since 2002 we have the Open Access Initiative – Protocol for Metadata Harvesting (OAIPMH) which is widely used by all kinds of institutional repositories archiving scientific material which most of the time fits right in to either two of the definitions of Digital Libraries, and if not they probably are functioning parts of a Digital Library as discussed above. The protocol provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAIPMH framework: Data Providers that administer systems which support the OAI-PMH as a means of exposing metadata; and Service Providers that use metadata harvested via the OAI-PMH as a basis for building value-added services. Data providers typically use repository software that are OAI compliant. Today there are several free software tools for creating institutional repositories available. Frequently downloaded, installed and used are programs such as E-prints, D-space, CDSware etc.

8.3 Building the Collection

One of the major issues of Digital Libraries at the moment is the question of creating a digital collection with some sort of scope and critical mass. This is especially true when it comes to offering full-text records. Building the collection usually means four things:

1. Acquiring original digital works created by original authors. This could be articles, books, conference proceedings, pictures etc.
2. Converting paper originals into digital format.
3. Purchasing or using free external material by either incorporating the material, such as electronic books and journals, or creating pointers to external websites.
4. Dealing with rights management issues, such as keeping track of copyright status of the digital material, identifying and authenticating users and their access to the material.

8.4 Document Formats and Retrieval

8.4.1 Document Formats

The Digital Library of today uses a wealth of different document formats storing and representing the content. A file format is a software algorithm for encoding the data and any information about the data. It can either be in a proprietary format, which means they are developed and used in software by commercial companies, or there are open file formats that are available for use by anyone. There are many hundred different file formats in use for different sorts of data. Talking about Digital Libraries and document formats. Here in this unit, we will only mention a few important formats which generally can be divided into different categories. One is Plain text formats where we find the most widely used character set encoding ASCII (American Standard Code for Information Interchange) (ISO 641), though it is fast becoming replaced by new and more sophisticated character sets such as UNICODE (ISO 10646-1) designed to cover all the worlds alphabets. Another is proprietary formatted-text formats including MSWord, WordPerfect and other word processing applications. There are also desktop publishing applications like PageMaker, QuarkExpress with their own format coding.

- **Rich Text Format (RTF)** is a export/import format developed by Microsoft and used by many word processers. Page Description formats describe shapes and the layout of a document. They are widely used for presentation of formatted pages in databases which carry journal articles in full development out of the Post Script (PS) programming language, developed in the Adobe laboratories in the mid 1980s. Structured information formats do not describe layout, but instead the structure of the

document. Standard Generalized Mark-up Language (SGML) and Extensive Mark-up Language (XML) are both meta languages which are used to specify an arbitrary arrangement of mark-up tags to be read by the computer. The arrangement of tags for different elements in the text makes it possible to create self descriptive objects which can be manipulated automatically by a computer system. The markup "grammar" is specified in a so called Document Type Definition (DTD). Different DTDs adapt to different sorts of documents. Encoded Archival Description (EAD) is much used for archival material; Text Encoding Initiative (TEI) is a DTD for humanities texts mainly. One of the most common used DTDs is Hyper Text Markup Language (HTML), which is a reduced tag set version of SGML.

- **Graphics Interchange Format (GIF)** for still images that store information about individual pixels or dots in the picture there are bitmapped formats such as Graphics Interchange Format (GIF). It is limited to 256 colours but is very popular because it displays well on computer screens and creates relatively small files. Portable Network Graphics format (PNG) is intended to replace GIF but has not yet been a public success. Joint Photographics Expert Group (JPEG) is on the other hand the most used compression format for colour photograph images on the web with over 16 million colour hues available. The Tagged-Image File Format (TIFF) is used as a archival or intermediary format. It stores large amounts of information and creates fields in the mega and gigabyte area. It supports all kinds of compression.

- **Computer Graphics Metafile (CGM)** for still images that use mathematical algorithms to store information about lines and curves, the standard vector format is Computer Graphics Metafile (CGM). For Audio and moving image formats the formats of choice lately is proprietary formats such as Audio Interchange File Format (AIFF) which is a Apple audio format. In the other corner we have an originally Microsoft Windows´ audio file format that now also works on other platforms (RIFF WAVE). RealAudio is another popular proprietary audio format which can "Stream" the information. That means the audio file starts to play as soon as the first bits are received by the users computer. Lately the audio format of choice has become MP3. It stands for MPEG Audio Layer 3. It is a compression

algorithm for digital audio developed by the Motion Picture Experts Group. MP3s are digital audio files that have been compressed while still maintaining good sound quality. This type of compression enables near stereo quality audio brought down to an extremely small size which made it so popular in the file sharing community. Multimedia collections of course also include moving image file formats. Many of the most used are proprietary like Apples QuickTime Movies, Windows Audio-Video Interleaved (AVI), Progressive Networks RealVideo and different versions of the compression standard for video – MPEG. In the multimedia classification we also find proprietory interactive formats like Macromedia´s ShockWave used for animated interfaces, advertisement and games production. Also found in multimedia documents is Sun Microsystems Java programming language or Microsoft´s Active X components.

8.4.2. Retrieval

Digital Libraries come in many forms. They can contain simple metadata or catalogues of bibliographical information. They can contain the full text of the document. They can contain images, audio or multimedia material. All this information may be available in different formats, created with different software. Most probably the resources reside on different servers using no unified thesauri or heterogeneous indexing schemes. All of this makes information retrieval a very complex task.

Every information system is unique when it comes to retrieval methods, and it is more or less necessary to have a fair idea of the characteristic features of each system to be able to perform a relevant search. This becomes even more complex when some Digital Libraries allow users to conduct searches across a range of distributed services. Those who work with human-computer interaction and usability testing have found that users-performance on computer-based systems can be greatly affected by the users´ previous experience and knowledge. This can generally be defined in one of two ways: the knowledge of the search topic or domain, and the knowledge of the system used.

Although it can easily be expected that the domain expert would outperform the novice with little knowledge of the domain, researchers point out that domain

specific knowledge begins to predict performance only after users have acquired some experience with the system used. In a study on factors affecting search performance on an on-line database system it was found that system knowledge and computer experience critically influenced search performance. It is concluded that the users search experience affected their use of search strategies and played a more important role than did their subject knowledge. So it seems that several studies confirm that both search experience and good search strategy are vital to get a good search result and that these variables mediate subject knowledge. Knowledge affects search strategy and search strategy in turn affects the researcher´s ability to locate relevant information. These findings highlight not only the importance of task-relevant knowledge but also the significance of the role that search strategy plays in the decision-making process. This knowledge has important implications for both designing more-effective information search aids and training the users of these aids. As the size of the Digital Library and complexity of issues increase, the nature of the search strategy is likely to become increasingly important in explaining the ability to locate relevant information. With very large databases, users are likely to be unable to utilize their knowledge to resolve issues unless they are able to limit the search to areas of the database where the information is located.

Web users, on the other hand, show very different patterns of searching from those found in traditional information retrieval systems such as online databases. For example, most users did not have many queries per search session, and each query tended to be short. Boolean operators were seldom used. Many users submitted only one query and did not follow with successive queries. So to support the typical web-searcher when he or she is entering a website that is using traditional IR principals one could recommend to improve the user interface by including broader resources such as increased browsing and viewing mechanisms and more online help-functions.

8.5 Preservation

When it comes to digital preservation and curation we are still living in some sort of digital stone-age. Many a task force on how to archive digital information for long-term preservation have come and gone. Usually their strategy recommendations are founded on infrastructures of common standards, methods

and tools. The big problem with preservation research is that it also must deliver practical results. This is often obstructed by reality where organisations use heterogeneous processes for archiving their material with structures that are incompatible for exchange and re-use of resources. But the need for standardized preservation technologies, practices and methods are very much in the public interest. The possibility to retrieve old as well as new cultural heritage material, research material, governmental documents etc. must be available to fortify democracy and to promote economic growth. One interesting effort to deal with this problem within Digital Libraries was recently made by a working group of the Network for Digital Libraries.

They focus on three research challenges that must be addressed if progress of sustainable Digital Libraries is to be made. The areas are:

- **Preservation strategies –** Emerging research domains that can identify new problems which are the result from constantly evolving technology. For example, we know a lot about media but lack archival media which is sustainable without intervention for more than 100 years.
- **Re-engineering preservation processes –** Today's preservation processes are slow, costly and often manual. Generally they follow the same roads as preservation practices for physical material. But if preservation processes shall become efficient a complete re-engineering must take place. Automatisation must be a key feature to speed up deliveries and slow down costs. Preservation functionality must be built into the systems used to create and manage the data. This is the only way to guarantee longevity of digital entities.
- **Preservation systems and technology –** To support this process transformation tools and technologies must be there and answer questions like: How can complex and dynamic entities be authenticated and their integrity verified? How can automatic mechanisms for the creation and authoring of metadata be accomplished? How can we solve issues of multilingualism for searching?

Preservation issues like these insist on answers and the answers must be delivered pretty soon or we will put a great deal of the worlds cultural and science heritage at risk.

8.6 Metadata and Naming

The reason why distances today seem to grow shorter and shorter could be spelled Interoperability. It is a key factor also for the Digital Library. The absolute trend of the 21st century is to adopt standards and develop and implement open systems that support interoperability. There are of course a multitude of aspects of interoperability but we will here only stop to take a short look on two of these aspects: metadata and naming. When we search the web we will find a lot of junk but the relevance in the OPAC-search is high. The reason for this fact is that the OPAC-documents are indexed based on a number of document-fields. These fields can be combined or be searched one by one, contrary to the web documents in the Google-search which contain a limited number of tags that can be indexed as fields. To incorporate value added fields into HTML-documents is of course unreasonable both from an economical and human perspective – but it would make life easier... There is nothing much that can be indexed as a field in an html-document – Title, URL, Link, heading...? Compare that with a MARC-index! Different persons have thought about this and arrived at one conclusion but offering different solutions. The conclusion is called metadata for web documents. This is simple information about the document which could be added by the creator himself in order to make the Web a better place for those who seek and those who want to be searched out. For this reason a number of different schemas for web document cataloguing have been created by librarians, archivists and researchers during the last 15 years. One of the more popular schemas is called Dublin Core. Dublin Core is a simple content description model for electronic resources. It got its name from the small town Dublin in Ohio, USA, where in 1995 representatives from museums, libraries, governments and commercial companies for the first time agreed on creating a core of metadata elements to improve search ability for electronic resources. The elements, it was agreed, should be so simple to use that the author of the documents would be able to catalogue his own resources. The metadata elements should contain a core of information used across most areas of science. The information in these fields must be enough for retrieval and identification of the resource. By using a common set of elements the Dublin Core group wanted to simplify the possibilities for semantic interoperability between different disciplines. Dublin Core has a broad international base for its work since about twenty countries in North America, Europe, Australia and Asia are involved in the developing work. The Dublin Core model is an economic alternative to more detailed description

models like MARC and other rules of cataloguing derived from the library world, but it is still sufficient and flexible enough to transpose structure and semantics from richer standards. The different need for metadata on the WWW calls for an infrastructure admitting coexistence of complementary and independent metadata schemas. Therefore the World Wide Web Consortium (WC3) has started to create a metadata-architecture called The Resource Description Framework (RDF) to support interoperability of metadata describing any item that can have a Uniform Resource Identifier (URI). In RDF Dublin Core, for example, is a standard vocabulary in the framework using XML as the encoding syntax.

Naming is a problem in the Digital Library. It is vital to have names that uniquely identify digital objects. These names must be part of any documents metadata. Names are important in order to give correct citations; make relevant information retrieval possible; to make links between objects and to manage copyright. Therefore naming must be permanent and can not be tied to a specific location. The name and the location must be separated. This is just the opposite to the current method used for identifying objects on the Internet today where we find information about the method by which a document is accessed. The machine name and the document path and the file which harbour the specific document is included in the same string in the Unique Resource Locator (URL). In this system when the file is moved the document is lost. That is why so much work has been done to find a scheme of unique identifiers that have persistence beyond the life of the server or the organisation. There are several of these schemes that have found solutions for giving documents persistent names that are valid whenever documents are moved from one location to another, which is often the case due to administrative reasons or migration from one medium to another.

Three major schemes have developed lately and they all are built around the idea to separate a document name from its location by mapping information of a unique never changing name to one or more URLs. They all are based on the assumption that an institution, like a national library, must take the responsibility for managing such a system. The most popular systems are Uniform Resource Name (URN) developed by the Internet Engineering Task Force (IETF); Digital Object Identifier (DOI) initiated by the Asssociation of American Publishers and the American Corporation for National Research

Initiatives; and Persistent Uniform Resource Locator system (PURL) developed by Online Computer Library Center in Ohio.

8.7 Digital Library Types and Content

The Digital Library community is clearly increasing in number and volume as more and more people get connected to high speed internet connections, more people get involved in distance learning, more people get used to online communication, governments, institutions and commercial companies realize the potential in digital deliveries. Developments like these have prepared the ground for a large number of different types of Digital Libraries throughout the world. It is difficult to classify a phenomenon with a definition still under
debate but for the purpose of this short introduction I would like to group Digital Libraries into 5 types and give a few examples of each:

- Digital Libraries at scientific societies or organisations
- Digital Libraries at Commercial publishers
- Digital Libraries at National Libraries
- Digital Libraries at Universities
- Digital Libraries at Museums and other cultural heritage organisations

8.7.1 Digital Libraries at Scientific Societies or Organisations

- **The Institute of Electrical and Electronics Engineers IEEE** provides access to almost one third of the world literature in the area of electrical engineering and computer science. Their Digital Library called IEEE Xplore provides full-text access to IEEE transactions, journals, magazines and conference proceedings published since 1988 and all current IEEE Standards. IEEE Xplore covers technical areas ranging from computer engineering, iomedical technology and telecommunications, to electric power, aerospace and consumer electronics. For full access you have to be a member (http://ieeexplore.ieee.org/)
- **Association for Computing Machinery (ACM)** members and registered users can use the Digital Library containing bibliographic information, abstracts, reviews, and the fulltext for articles published in ACM periodicals (journals, magazines and transactions) and ACM

proceedings (http://portal.acm.org/). Close to one million documents of interest to people working in particle physics and related areas can be found at the CERN document server site. Originally named Conseil Européen pour la Recherche Nucléaire, now renamed European Organization for Nuclear Research, the Digital Library at CERN covers preprints, articles, books, journals, photographs, and much more available at no cost for everyone (http://cdsweb.cern.ch/).

- **BioMed Central** is an independent publishing house committed to provide immediate free access to peer-reviewed biomedical research. BioMed Central publishes more than 50 online journals covering the whole of biology and medicine. The service includes support for journal editors in developing countries (http://www.biomedcentral.com/).

8.7.2 Digital Libraries at Commercial Publishers

The content of these Digital Libraries are mostly the same as of Digital Libraries of scientific societies and organisations – bibliographical or full text copies of journal articles, conference proceedings etc. drawn from a single or distributed databases. Major examples of this type are:

- **Springer Link,** with a collection of journals and book series that account for over 1 million documents in different kind of subject areas from Springer Science Publishers (http://www.springerlink.com/).
- **ScienceDirect** which is the giant publisher Elseviers collection of science, technology and medicine full text and bibliographic information of the same kind as you find in Springer Link (http://www.sciencedirect.com/).
- **ISI-Web of Knowledge** accesses multidisciplinary databases of bibliographic information gathered from thousands of scholarly journals. The databases are indexed so you can search for specific articles by subject, author, journal, and/or author address. Because the information stored about each article includes the article's cited reference list you can also search the databases for articles that cite a known author or work. (http://portal.isiknowledge.com/)

8.7.3 Digital Libraries at National Libraries

National libraries are the collective memories of nations. They always house valuable collections of both scientific and cultural nature where citizens go for research and investigation of historical heritage and events. These days many national libraries take advantage of modern technology to help them serve their customers need. By digitizing text, sound, film-collections and making them into Digital Libraries within the national library, cultural heritage treasures are made available for people far outside the national
borders. Fine examples of this are to found at:

- **British Library** which offers a number of digital information services based on British Library collections. One example is the "Treasures in Full" site that brings high-quality editions of the works of Shakespeare, Chaucer and the Gutenberg bible among other things to the users desktop (http://www.bl.uk/treasures/treasuresinfull.html). Another is "Images online" which gives access to thousands of pictures from the library collections (http://www.imagesonline.bl.uk/britishlibrary/).
- **Library of Congress** offers two great Digital Library resources: "American Memory" which is an umbrella term for a collection of digital resources on topics such as African American History, Immigration, Native American History, Performing Arts etc. (http://memory.loc.gov/ammem/) and a resource called "Thomas", named after Thomas Jefferson, made up of several databases, where federal legislative information is freely available to the Internet public (http://thomas.loc.gov/).
- **The National library of Portugal** can stand as an example of a smaller national library with limited resources that is working towards a vision of a National Digital Library consisting of a coherent group of services and resources with technical solutions based on open and scalable technology. Several digital collections and exhibitions are available. A good example is the digital collection devoted to the renaissance portugeese mathematician and astronomer Pedro Nunes (http://purl.pt/40/1/)

8.7.4 Digital Libraries at Universities

In the second half of the 1990s several university libraries started building digital collections making them public available.

- **The Electronic Text Center at the University of Virginia** is one famous Digital Library offering thousands of SGML-encoded electronic texts and

many special collections devoted to famous authors or American historical events (http://etext.lib.virginia.edu/).

- **Project Gutenberg** also emanated from a university. In this case the University of Illinois. The objective of the project was to provide free access to digital version of world literature. The texts are stored in plain ASCII format making them easy to read and search with any sort of computer equipment. Today Project Gutenberg is a volunteer effort continued outside the university (http://www.gutenberg.org/).
- At the **Oxford Digital Library** a number of disparate collections from the university are available. You can go from images of medieval manuscripts to a database of Athenian Pottery to a collection of motoring and transport images. In the future the Oxford Digital Library aims to offer a Digital Library architecture which will allow centralized access to these digital resources. The use of established standards for descriptive metadata (i.e. EAD, TEI) is a precondition for this integration process. Existing Digital Library collections may be transferred step by step into a common architecture with an integrated retrieval mechanism (http://www.odl.ox.ac.uk/).
- Established in 1997 as a **University of California library**, the California Digital Library has become one of the largest Digital Libraries in the world providing access to resources like The Online Archive of California – collections of digital materials (such as manuscripts, photographs, and art) held in the libraries, museums, and archives across California; government data and statistics about California in "Counting California"; The Melvyl catalogue with its 15 million records from the 10 University of California campuses. (http://www.cdlib.org/).

8.7.5 Digital Libraries at Museums and other Cultural Heritage Organisations

Many cultural heritage institutions are building digital collections of their holdings in order to provide easy and affordable access to the cultural heritage resources. In many ways it has been a tough trip since funding and research for Digital Libraries

of this type have not been as excessive as in other areas. There is also the problem of common standards in order to describe cultural objects homogenously, using the same kind of metadata standards. Cultural heritage institutions use a multitude of different standards and in many cases no standards at all. This will be one of the big challenges for the next generation of Digital Libraries at museums and cultural heritage sites.

- The **State Hermitage Museum in St. Petersburg, Russia** has provided access to many collections of the museum. A variety of techniques have been used for making 3D images and virtual exhibitions come alive in both Russian and English (http://www.hermitagemuseum.org/)
- The 24 Hour Museum is the **UK's National Virtual Museum**, offering a unique mix of dynamic content including daily arts and museum news as well as exhibition reviews. It functions as an access point to Cultural heritage sites of the UK. Venue and listings info is driven by a comprehensive searchable database of more than 3,400 entry points. This brief expose hopefully has made clear how varied and heterogeneous the flora of Digital Libraries are today. Some work with very simple technology, like Project Gutenberg, but others like virtual worlds of the Hermitage are very sophisticated. Some are designed to provide access to digital resources in specific fields like BioMed Central. Others give access to specific document types in a wide area of scientific subjects provided by commercial publishers that charge you for the data, while many other Open Access archives are available for free.

8.9 Conclusion

As we have seen Digital Libraries are thriving and expanding their services worldwide not caring too much if they fit into one definition of some sort or the other. Many have been sponsored and funded by government bodies such as the eLib Programme in the UK and the Digital Library Initiatives phase 1 and 2 in the United States. What is being done at university libraries, at commercial publishers,

scientific societies, museums or any other organisation producing and managing information for a community of users, is much more than converting analog data into digital form. New material is created and served in new forms. Take for example how resources from different collections are put together into new entities of information by making great use of pictures and sound and all kinds of different search facilities, which is just impossible, for economical or technological reasons, to bring about in the analogue world. These resources go side by side with full-text journals, books, references available at your desktop in seconds. There are several stumbling blocks that still irritate the everyday user and the visionary when using the Digital Libraries of today. We have, for example, the technical issues which include the problem with standards and protocols. To bring the distributed variety of digital resources and services together in a way that allow for integration and unified search, retrieval and presentation is a great challenge for the future. So is the problem of transferring personalised service and support from standard library and information services to the Digital Library. A user interface can hardly replace person to person service but better user interfaces must be developed and researched in order to help users. The future Digital Library will go beyond helping the user with searching and browsing only. Users must be able to expect support for taking correct actions and getting help for problem solving where the Digital Library system confirm or deny existing hypotheses, distinguishes between traditional searching and browsing which is called "tactical level cognition" and the problem solving act which is called "strategic level cognition". In the future, Digital Libraries must become not only a simple storage place but a place where knowledge is acquired, shared and multiplied. To facilitate the browsing function Digital Libraries must integrate diverse repositories of coherent collections and include navigation, searching and browsing facilities in a network of inter-related concepts and repositories.

Self-assessment Questions

1. Briefly explain the digital library development.
2. Discuss the various formats of digital contents, their preservation and retrieval.
3. Explain with examples the different types of libraries available online and their collection in detail.

Activity:

1. Visit the website of IEEE and prepare a comprehensive report on its resources, services, and salient features.

Recommended Readings:

1. Agarwal, P. K. (1999). India's national internet backbone. *Communications of the ACM*, *42*(6), 53–58. doi:10.1145/303849.303862

2. Agnew, G. (2008). *Digital Rights Management: A Librarian's Guide to Technology and Practice*. Oxford: Chandos.

3. Ahmad, R. (2002, December). Forging a Sustainable Development Model: The Malaysian way. *Development*, *45*(4), 74–79. doi:10.1057/palgrave.development. 1110409

4. Ahsanullah, A. (2002). *Information Technology Services for Rural Communities of Developing Countries toward poverty Alleviation*. Dhaka: Faruque & Aulad Hassain.

5. Arms, W. Y. (2000). Automated digital libraries: how effectively can computers be used for the skilled tasks of professional librarianship? *D-Lib Magazine*, *6*(7/8). Retrieved from http://www.dlib.org/dlib/july00/ arms/07arms.html. doi:10.1045/july2000-arms

6. Arms, W. Y. (2001). *Digital libraries*. Cambridge, MA: MIT Press.

7. Arms, W. Y., Blanchi, C., & Overly, E. A. (1997). An Architecture for information in Digital libraries. *D-Lib Magazine, 3*(2). Retrieved April 28, 2009, from http://www.dlib.org/dlib/february97/cnri/02arms1.html*Art Museum Image Consortium*. (n.d.). Retrieved from http://www.amico.org/

8. Atkinson, R. J. (1997). Towards a more secure internet. *IEEE Computers*, *30*(1), 57–61.

9. Baker, D., & Evans, W. (2008). *Digital Library Economics: Digital library economics: An academic perspective*. Oxford, UK: Chandos Publishing.

10. Chowdhury, G. G., & Chowdhury, S. (2001b). *Information Sources and Searching on the World Wide Web*. London: Library Association Publishing.

11. Chowdhury, G. G., & Chowdhury, S. (2003). *Introduction to Digital Libraries*. London: Facet Publishing.

12. *Digital Preservation and Storage*. (2004). Technical Advisory Service for Images.

13. Gillespie, J., Fair, P., Lawrence, A., & Vaile, D. (2004). Coping when Everything is Digital? *Digital Documents and Issues in Document Retention – White Paper*. Baker & McKenzie 2005 from http://www.bakercyberlawcentre.org/ddr/

14. World Development Indicators. (2008). Washington, DC: World Bank.

Unit. 9

DIGITAL LIBRARY DEVELOPMENT: SOFTWARE

Complied by: Muhammad Jawwad

Reviewed by: Dr. Amjid Ahmed

CONTENTS

INTRODUCTION

The unit is specifically designed for the students and librarians who are interested to develop digital libraries. Some important digital library software/system was discussed in the unit in detail. Benefits and important features of digital library systems are also elaborated. The unit will outline design and architecture models of digital library system. The special focus has been given on the overview of open-source digital library system particularly DSpace, Greenstone and EPrints.

OBJECTIVES

After studying this unit, the students will be able to explain the following:

1. What is digital library software or system?
2. What benefits and features a digital library software possess?
3. What is the design and
4. architecture of digital library system?
5. How many open-source digital library management systems are presently available and what are the important features and limitations of these systems?

9.1 Introduction

The design of systems for management and delivery of digital library content is an interdisciplinary area where research on digital libraries intersects with software development, database management, information retrieval, and human-computer interaction. Digital library management systems (DLMS) share some similarities with web content management systems but are also different because of the required support for digital library standards, especially in regard to information organization and interoperability. DLMS represent a specialized category of software systems that integrate functionality for building, managing, storing, providing access to, and preserving digital objects and collections. They are part of a broader category of digital asset management systems that are used in practice for acquisition, indexing, storage, management, preservation, and delivery of digital objects. In a distributed digital library environment, DLMS also provide platforms for aggregating digital content and metadata.

The design of DLMS has been an area of active development since the early days of digital libraries and has included efforts to provide conceptual models as well as platform solutions. Many of the early systems were custom-built, designed for single projects in order to meet the needs of a particular community. The late 1990s saw the development of the first architectural models for repositories and the emergence of the dedicated content management systems for cultural heritage digital collections. Greenstone was released as open source software and has been widely adopted throughout the world. It became a popular choice in the category of proprietary software among the US public and academic libraries. Early 2000s marked the construction of the first digital repositories for scholarly publication with EPrints developed at the University of Southampton, UK, and DSpace and Fedora released in the United States. Currently, there are number of open source systems and commercial software packages available for building digital library systems and meeting the needs and requirements of specific communities. Many organizations, however, especially those building large-scale systems such as Europeana, HathiTrust, or National Science Digital Library (NSDL) have developed custom platforms. In recent years, libraries and archives have been migrating from the first generation of open source or proprietary software, such as DSpace toward more robust and scalable open source solutions. The new

generation of DLMS is comprised of several open source technologies and often integrate Fedora with other customizable platforms, such as Hydra and Islandora.

9.2 What is Digital Library Software?

Digital library software is a software helps in digitally organizing library items like books, journals, catalog, etc. It provides easy and affordable access to the library. The purpose of the digital library is to underpin learning and acquisition of knowledge, to provide a solid base for education and enhance quality of life. Digital library software is based upon two types of technology; bar code system and RFID system. The bar code system is based upon bar-coding and RFID is based upon the RFID tags and RFID scanner. As we all know that it has reduced the workload of the librarian by systematically organizing the database and library items in digital form. But they require abreast of themselves with the technology used in the library management system. In the absence, of knowledge, they will face many technical problems like updating the database, generate records, etc.

9.2.1 Definition of Digital Library Software/System (DLS)

The terminology of DLMS and DLS is used according to the DELOS Manifesto. However, it is important to note that other terms are also used for DLMS, including digital content management systems, digital collection management systems, repository platforms, or digital asset management systems. The term “digital asset management system” is being used widely and often interchangeably with DLMS. In practice, its usage is broad and often encompasses a wide range of software, including digital collection systems (e.g., CONTENTdm, Omeka), repositories (e.g., DSpace, Fedora, Digital Commons), digital preservation systems, or even databases, such as FileMaker Pro.

The concept of DLMS here we use as according to the definition proposed in the DELOS Manifesto as “a generic software system that provides the appropriate software infrastructure both to produce and administer a Digital Library System incorporating the suite of functionality considered foundational for Digital Libraries and to integrate additional software offering more refined, specialized, or advanced functionality”.

The DELOS Manifesto makes a distinction between a digital library system (DLS) and a DLMS. A DLS offers functionality for a particular digital library, including support for end user interactions. A DLMS provides a platform for producing and administering digital collections and services by ensuring essential functionality and incorporating additional software components for more refined and advanced features. DLMS enable an instantiation and management of digital collections and services that become part of a centralized or distributed DLS.

9.3 Benefits and Features of DLS

The digital library software is taking a boom with the advancement of technology. It has become necessary for schools, universities, institutes etc. to meet the current need of the modern library. Modern libraries help in enhancing the skills of the students as they can explore more books with the availability of online records and save their time in searching for a book. The digital library system works upon LAN networking. Without a local area network, the digital library system won't work and generate results. Networking involves sharing of computers and linking them to communication channels. In networking computers play an important part of digital system. In Library Management System the network is also called the tree structure in which each segment is sub-divided into other nodes. The storage efficiency of network structure are commonly used with mainframe and minicomputer systems. Digital Library Management Software is a computer-based system where all the functions of the library are controlled digitally and in a systematic way. It helps in keeping a record of books, journals, catalogs, newspapers, CDs, etc. in the library. The benefits of the digital library Management Software are:

- Easily accessible
- Track any information at any time
- Act as an anti-theft equipment
- Maintenance of huge database
- They are an active system of learning and knowledge
- Reduce operating cost

Library Management System is also known as an integrated library system that manages all the activities such as acquisition, circulation, stock verification,

cataloging, etc. Library Management System Software helps the students to remain up-to-date with the library database and other resources.

Digital Library System is an Advance Library Software that helps in fast and effortless gathering of all information in the library. As we all know the library is an important part of all schools and universities and the demand for Library Automation Software has increased with the advancement of the system. We can make our library a smart library with Library Administration Software which will turn our library into a digital system where we can easily access the database. As the digital revolution has already taken a boom, every school, college and university also wants to get advance and remain up-to-date with the current market demand. The library system has taken gear due to Government educational policies and systems. The government provides funding to the schools, colleges and universities for the installation of the Library Automation and Digitization System.

Following are the features of digital library software:

- It provides speedy and wide access to information
- It saves time
- It works as an anti-theft
- Systematic order of books
- Generate reports
- Easily accessible
- Information can be retrieved
- The database is maintained by all users

9.4 Design and Architecture of Digital Library System

Digital Library System serve diverse groups of users from scientific, educational, archives, and museum communities. They provide access to scholarly publications, data sets, archival documents, and cultural heritage objects in multiple formats and languages that are described by a variety of metadata standards. Obviously, there is no single, universal software system that could meet the needs of all user communities and support the variety of data types and metadata schemas. In

addition to user requirements and functionality, the design of a DLMS has to address the technical aspects of system reliability, scalability, and sustainability.

Most digital libraries are created independently by content holders in research, library, archives, and museum communities using a range of standards and software solutions. Interoperability has become one of the most important issues in the development of digital libraries. The goal of interoperability is to enable the exchange of data between independent digital libraries and to provide services for easier discovery and interaction with digital library resources in the network environment The initial systems were often built "from scratch" or incorporated existing software components but offered limited modularity and interoperability.

Early DLMS were standalone systems and had typical features of the system-centered design, which meant they were difficult to install, customize, or configure. The large-scale digital libraries that have emerged in recent years pose new challenges for interoperability and system reliability, scalability, and sustainability. The aggregation of content and services can take place on consortial, regional, national, or international levels. These large-scale digital library systems are built either as centralized aggregators of content, metadata, and services or use a distributed network of content and service hubs with a service layer to facilitate access to harvested metadata and links to objects. HathiTrust is an example of a centralized model, while Europeana, the Digital Library of America, and the National Science Digital Library represent large-scale distributed systems. The first two decades of digital library research efforts have concentrated on defining the components of digital library systems, delineating the relationships among them, and developing conceptual models that would enable interoperation between individual DLS.

This unit provides an overview of the research surrounding digital library architectural models, reviews functionality and other system requirements, and discusses interoperability approaches.

9.4.1 Architectural Models of DLS

The concept of architecture in the context of digital library systems refers to "a consistent and comprehensive set of software components necessary for a DLS and the interaction between them". An architectural model serves as a starting point and

a conceptual framework for designing flexible and interoperable systems. It is an abstract framework for identifying components and describing the relationships among them, usually independent of specific standards, technology, and implementations. The goal of an architectural model is to provide a set of common basic elements and to serve as a blueprint for an integration of sub elements supporting specialized functionality. The researchers and experts' community has proposed several models of digital library architecture of various levels of complexity. Three core components can be identified across the frameworks, although the names of the components often vary:

1. Data store (also referred to as content files, data repository, or archive)
2. Metadata registry (also referred to as metadata catalog, metadata store, or index)
3. A set of services (e.g., index, store, manage, copy, authorize, etc.)

9.4.2 Functionality and other System Requirements

Designing DLMS is an extremely challenging task, as it requires the integration of architectural models, technologies, and standards. It begins with a conceptual model but also involves a range of technologies, standards, and applications. It is a combination of those elements that contributes to a flexible and usable system design and provides the required functionality for creating, managing, and using a digital library. Functionality refers to system's capabilities in building and managing digital libraries and providing end user support. Functionality is expressed in terms of services and is often divided into fundamental and value-added services. The set of core functions relates to object- and collection-building, managing, disseminating, and/or preservation capabilities.

Researchers sometimes distinguish preservation services from those focused on content creation and management, but since many systems integrate preservation and access services, the following list includes preservation in the core functions:

1. Creation of digital objects and collections, which includes ingesting and/or processing of digitally-born or digitized materials and associated metadata records; creation of collections of objects based on predefined selection criteria

2. Management, which involves adding, modifying, and deleting objects; management of user rights and permissions
3. Access, which includes indexing, searching, browsing, and harvesting services plus presentation of objects and collections through a web interface and tools for user interaction • Preservation, which includes services to store and manage digital objects and archival master files.

In addition to core functions, DLMS can provide a wide range of advanced and/or value-added services. Some of the services, although not included in the core, are critical to building digital libraries and maintaining them over time. For example, Export service is a function of the system that provides a means of retrieving objects and/or metadata and depositing them outside of the system. Export function becomes critical when institutions decide to migrate to a different platform. There is no comprehensive list of digital library services, although some efforts have been undertaken to define services based on theoretical models.

9.4.3 Interoperability

Interoperability refers to a system's ability to communicate with other digital library systems using standard protocols in order to exchange data. Interoperability has many aspects including uniform naming, metadata formats, document models, and access protocols. It has been recognized as a critical problem and a fundamental challenge since the early days of digital library development. The goal of interoperability is to build a set of services for users "from components that are technically different and managed by different organizations". The challenge lies in heterogeneous content, multiple data formats, different protocols, and the variety of metadata schemas used by individual organizations. Establishing an interoperability framework is not only a technical but also organizational issue, as it requires a variety of content providers to cooperate and agree on common standards.

1. Three basic approaches have been identified in the "spectrum of interoperability" with different levels of engagement from content providers: • Federation provides the strongest form of interoperability, but it also places the highest demands on the participating institutions. It requires that content providers agree that their services will conform to certain specifications. Federation is a well-established form of

exchanging data in the library world. Examples of federated services include the sharing of online catalog records using Z39.50 protocol or metasearching of multiple journal databases.

2. Harvesting represents a less rigorous approach. Participating institutions have to agree to expose and share their data, but they don't have to adopt a formal set of agreements.
3. Gathering represents the least demanding approach for content providers. Resources openly available on the web are gathered by web crawlers, and no formal agreement may be necessary between organizations holding digital content and a digital library service provider collecting it.

Digital library service providers may select one of the approaches or a combination thereof in an effort to aggregate metadata and/or content from multiple independently operated libraries or other content providers.

9.5An Overview of Open Source DLMS

The Open Source Digital Library software are the software applications that help in creating and presenting information repositories. The repositories built with the help of these Digital library management systems can be searched and browsed based on Metadata as these features are inbuilt in such applications. Apart from this, they can be easily maintained, enhanced and re-created.

Presently many open source software (OSS) applications are available for library and information management, for example DSpace, GSDL, Fedora, Eprints etc. Therefore, organisations can choose the one which is the most suitable for their requirement and implement them to create digital repositories. Here in this unit, we focused mainly on three of the most popular Open source Digital Library software- DSpace, GSDL and EPrints.

- **DSpace** DSpace is an open source digital library software which allows us to capture and store digital data like text, video, audio etc into created repositories. It also provides facility to index, preserve and disseminate the digital material. Thus digital libraries use DSpace to manage the digital materials and publications in professionally maintained repositories. If we

see the world-wide scenario, there are more than 1000 digital repositories which are developed using the DSpace application for storing, distributing and preserving their digital data. DSpace is more common as a platform to build an institutional repository which is a digital collection of research documentation, intellectual publications, library collections etc. In Indian scenario Dspace is being used in many reputed organisations and projects like National Digital Library Programme of GoI, IIT Kharagpur Central Library, DIAT, DU (Deemed University) Pune, KUVEMPU University other IITs, IIMs and many other research and academic organizations.

DSpace performs three major tasks to build a repository:

1. It captures and ingests the digital content along with metadata
2. It lists the content systematically and helps in searching based on keywords and metadata.
3. It supports preservation of the digital data for a long period of time.

Therefore, DSpace can easily be customised to manage and preserve the digital content and provide accessibility of this data to the users. Since it is an open-source software, an active community of developers, researchers and users across the world are collaborating to provide their expertise to enhance this application. DSpace is capable of storing a wide range of digital data, which includes documents like articles, technical reports, conference papers, books, theses, multimedia publications, administrative records, images, audio-video files, web pages etc. It also provides multiple features like visualization, simulation of the stored data etc.

A) **Latest Features of Dspace**: As DSpace is a continuously growing platform, it keeps on releasing upgraded versions from time to time. 6.x is the latest update to the DSpace platform. It consists of an upgraded configuration system, upgraded file storage plugins, and better-quality control / health-check reporting features (through REST API and also through email). Furthermore, DSpace has a Java API refactor that adds support for both UUIDs and Hibernate in the database layer. This feature makes it compatible for future challenges. As reported by

DSpace official website, the new Features and improvements in 6.x version includes.

- Java API refactor, featuring Hibernate and UUIDs
- Enhanced (reloadable) configuration system, featuring a new local.cfg configuration file
- Enhanced file storage plugins, featuring support for Amazon S3
- Configurable site health checks via email
- XMLUI framework for metadata import from external sources, featuring support for PubMed imports
- XMLUI export of search results to CSV (for batch editing) • XMLUI extensible administrative control panel
- REST API Quality Control Reports, along with sample HTML clients and CSV export (for batch editing) • REST API support for additional authentication methods (e.g. LDAP, etc)
- All searches default to Boolean AND.
- Enhanced indexing for searches (Excel is now searchable, as well as right-to-left text in PDFs)
- OAI-PMH adds compliance for Open AIRE 3.0 guidelines for literature repositories"

B) Limitation of Dspace: During implementation some limitations have been observed such as Flat File and Metadata structure, poor user interface, lack of scalability and extensibility, Limited API, Limited Metadata Features, Limited Reporting Capabilities and lack of support for linked data.

- **GreenStone Digital Library Greenstone Digital Library (GSDL)** is an open source, multilingual software, which has been released under the terms of the GNU General Public License and is used widely for creating repositories and making them accessible online. The development and distribution of GSDL is an outcome of the joint efforts by the New Sealand Digital Library Project at the University of Waikato, UNESCO and the Human Info hyperlink "http://humaninfo.org/" NGO.

The aim of Greenstone software is to enable the users in building their own digital libraries. It provides a way to organise this information and publish it on the web or any other digital storage media like DVD and USB flash drives. In the later case, it will run on a non-networked environment. The digital libraries built by GSDL are fully-searchable and metadata-driven digital resource14. Infact, this software encourages the effective deployment of digital libraries to share information and put it in the public domain. Therefore, it is in itself not a digital library, rather it provides a platform to build the digital library.

In 2004 its developers of GSDL were awarded by IFIP Namur award for "contributions to the awareness of social implications of information technology, and the need for a holistic approach in the use of information technology that takes account of social implications".

GreenStone Digital Library Versions There are two main versions of GSDL namely GSDL2 and GSDL3. GSDL2 was the earlier version and still under wide-use where as GSDL3 is the latest version under active development. The best thing is that GSDL3 has backward compatibility and contains almost all the features of GSDL2. If a programmer is already working on GSDL2, he can either work with the latest release of GSDL2 or consider upgrading to GSDL3.

The **Greenstone Librarian Interface (GLI) provides a feature** to import 'Greenstone2 collection' which helps in migrating to the new software for existing users of GSDL2. Greenstone3 has been developed in JAVA and uses various latest web technologies—like XML Transforms (XSLT), and the Java Authentication and Authorization Service (JASS). In the same context if we see Greenstone2, then it was written in C++ and was based on many self-developed techniques by the developers as many latest web technologies were not available at the time. This made the users totally dependent upon the documentation by the development team. All these limitations have been overcome in the latest GSDL version.

Limitation of Green Stone Digital Library Some limitations of GSDL have also been observed like Interactive content updation and management are not possible, no provisions for identifying duplicacy, metadata handling seems to be a bit difficult, during the collection building processing of some documents it hangs. Also, Linux Version looks robust than Windows.

- **EPrints Eprints** has been one of the popular Digital library software which has been in use for almost last two decades It has been created at the University of Southampton and the currently version EPrints 3.3.16 Beta 1 is being used. Being an open source software, it is convenient for use by any organization with limited resources also.

  Initially Eprints required software-platform like Linux, Apache, MySQL, and Perl; now it can also run on Window's platform which has made it even easier for users. Just like the other two Digital library software, Eprints is also a good choice to create an Institutional Repository and make it running. Documents along with the necessary metadata for the records can be uploaded by the users by filling information into a web form. This software links to the SHERPA/RoMEO database which helps the authors to verify their rights regarding their submissions in the repository. In this way any unauthorized submission by the content-publisher is well taken care of.

  Features of Eprints: Eprintsis easy to use for both the end-users and the administrators; this is the biggest quality of Eprints. Users can submit the documents on Eprint in a straight-forward manner where users can proceed through the submission-process one step at a time. The metadata information can be provided with the e-copy of the document. The metadata information is quite simple like document type, document-title, author's name, date of submission

etc. and can be submitted using a simple form. This doesn't require any knowledge of HTML or XML. For the administrator, the fields in the metadata are customizable. Therefore, the administrator can allow only those fields which are relevant for a particular repository and the end-user needs to fill only those particular fields. Users have an added advantage to manage their submissions as editing, updating, and removal of documents is possible even after submission. However, the administrator has the rights to restrict these functionalities.

Another facility that Eprints provide is that the administrator can specify a period only after which the document is transferred automatically to the archive-section. Eprints also provide very effective search as well as browsing features. Search can be performed based on multiple options whereas the browsing feature is customizable and robust. This helps in finding the documents effectively in the archives ("Repositories Support Project"). The Metadata Field entered, help in browsing the collection. For example, a particular document can be browsed Year-wise, departmentwise, volume-wise etc. Browsing can be done based on any of the metadata fields within a collection, and multiple browsing criteria can be used. The browsing category can be customised by the administrator. Since Eprints is OAI-compliant, Google indexes the documents which are uploaded on an Eprintsarchive. This helps in enhancing the visibility of Eprint documents in cyber-space. As per the feedback provided by users and other technical reviews, it has been widely accepted that the installation and configuration of Eprints is simple and fast. 'Eprints Services' is a company formed by the developers of Eprints which helps organisations to install, configure and use Eprint based repositories.

Due to its multiple advantages today Eprints is being used in approximately 300 reputed organisation, the largest being the repository developed at the University of Twente in the Netherlands. This repository contains over 60,000 record. This in itself demonstrates the capability of Eprints in handling large collections.

Limitations of EPrints: No doubt there are multiple advantages of using Eprints to create digital repositories in libraries; still we may count certain limitations like the lack of the bulk upload feature. Uploading of files and creating records is definitely easy, but if someone has to upload an existing archive, then there are no options available to upload multiple records at one time. Multiple files can be uploaded in one go, but only when belong to the same record. To elaborate further, migrating of records from an existing digital library software to Eprints is not at all a problem but if the existing collections are not contained within a database, then the records can't be uploaded in bulk in Eprints. This means each record has to be created individually. Also, in Eprints one can't create common records for multiple documents rather individual records for each document should be created one by one. Another limitation of Eprints is the limited features in its search functionality. Boolean search is not available and also sometimes the search gives no output at all, which is not acceptable in today's time. At least suggestions for alternate search should be provided. User-created tagging feature is also missing in Eprints.

9.6 Conclusion

The design of DLMS has evolved in the past decade with the introduction of low-barrier systems, robust and flexible open source repository systems, and more specialized options available to specific user communities like archives and museums. Selection of an appropriate system among so many alternatives,

however, is not easy. The variety of DLMS and their features makes the selection and evaluation process challenging. In addition to the “out-of-the-box” solutions, institutions have the option of customizing available open source software or building their own custom systems. Organizational requirements depend on the types and characteristics of resources, needs of intended users, traditions of information organization and resource sharing, strategic goals, consortia agreements, and the technical infrastructure of individual institutions.

There are four major areas are identified for functional requirements: information organization focusing on content and its associated metadata, presentation, access including interfaces for both internal and external users, and preservation. The landscape of DLMS has evolved from the early standalone systems to flexible, open models and collaborative, multipurpose systems. The development of open source software and the competing systems from commercial vendors provide digital library developers with many options to create and manage digital content.

The increasing diversity of DLMS does not mean that the currently available systems meet all user requirements. While there has been significant progress in interoperability and building flexible systems, user-level functionality still leaves much to be desired. DLMS function for the most part as databases of objects and associated metadata with limited capabilities for building layers to present contextual information. Further, they are limited in their ability to provide workspaces for end users to manipulate digital objects, contribute their own materials, or collaborate with others.

Number of hypothetical scenarios for advanced digital library functionality to support users in cultural heritage and scientific communities have been discussed by researchers and experts. Many concepts proposed in those scenarios, such as real-time construction of collections, personal annotations,

or collaborative spaces, are still not supported by most standard DLMS. The design of DLMS is still a work in progress.

Self-assessment Questions

1. Define digital library management system (DLMS) and describe its features and benefits.
2. Explain the design and architecture of DLMS.
3. Why open-source digital library software are popular in the library community? Explain through examples.
4. Write comprehensive note on the DSpace DLMS.

Activity:

1. Prepare a comparison chart of Open-source DLMS described in the unit.

Recommended reading:

1. S Andro, M., Asselin, E., Maisonneuve, M., 2012. Digital libraries: comparison of 10 software. Lib. Collect., Acquisit. Tech. Serv. 36 (3), 79–83.

2. Arms, W.Y., 1995. Key concepts in the architecture of the digital library. D-Lib Mag. 1 (1).

3. Arms, W.Y., 2000. Digital Libraries. MIT Press, Cambridge, MA.

4. Arms, W.Y., Hillmann, D., Lagoze, C., Krafft, D., Marisa, R., Saylor, J., Van de Sompel, H., 2002. A spectrum of interoperability: the site for science prototype for the NSDL. D-Lib Mag. 8 (1), 9. Available from: http://www.dlib. org/dlib/january02/arms/01arms.html.

5. Candela, L., Castelli, D., Pagano, P., 2007a. A reference architecture for digital library systems: principles and applications. In: Digital Libraries: Research and Development, pp. 22–35. Berlin: Springer.

6. Candela, L., Castelli, D., Pagano, P., Thano, C., Ioannidis, Y., Koutrika, G., Schuldt, H., 2007b. Setting the foundations of digital libraries: the DELOS

manifesto. D-Lib Mag. 13 (3), 4. Available from: http://www.dlib.org/dlib/march07/castelli/03castelli.html.

7. DSpace 2015, DSpace Home. Available from: http://www.dspace.org/.

8. DuraSpace, 2015a. Fedora user registry. Available from: http://www.duraspace.org/registry/fedora.

9. DuraSpace, 2015b. Hydra community framework. Available from: https://wiki.duraspace.org/display/hydra/Hydra +Community+Framework.

10. Suleman, H., 2005. Analysis and evaluation of service-oriented architectures for digital libraries. In: Peerto-Peer, Grid, and Service-Orientation in Digital Library Architectures, Springer, Berlin and Heidelberg, pp. 130–146.

11. WGBH Educational Foundation, 2015. Open Vault. Available from: http://openvault.wgbh.org/.

12. Witten, I.H., 2008. The development and usage of the Greenstone digital library software. Bull. Am. Soc. Inf. Sci. Technol. 35 (2), 31–38.

13. Witten, I.H., Bainbridge, D., 2003. How to Build a Digital Library. Morgan Kaufmann, San Francisco.

14. Yeh, J.H., Sie, S.H., Chen, C.C., 2009. Extensible digital library service platform. In: Theng, Y.-L., Foo, S., Goh, D., Na, J.-C. (Eds.), Handbook of Research on Digital Libraries: Design, Development, and Impact. Information Science Reference, Hershey, NY, pp. 27–40.

15. Zhang, A.B., Gourley, D., 2009. Creating Digital Collections: a Practical Guide. Oxford, England.