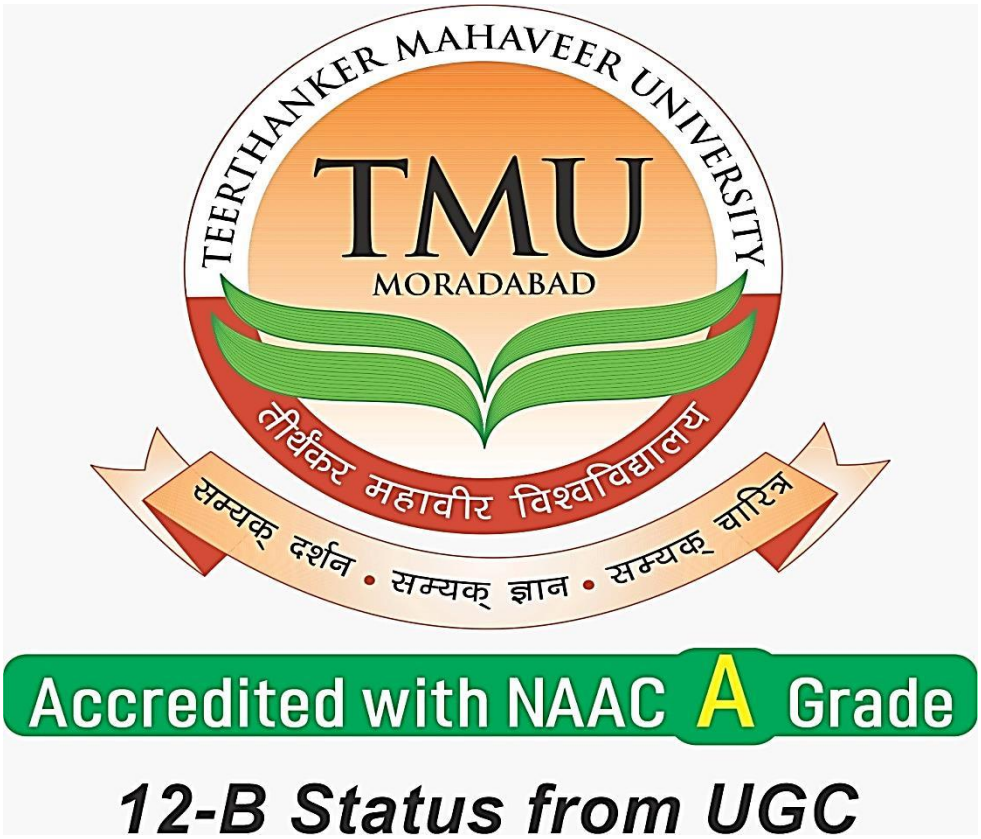


Digital Libraries

DBLIS202

UNIT 1
Introduction to Digital Library

1. Introduction

The computerization of the library during past few decades have focused heavily on the creation of surrogate records of printed documents available in a library or for providing computerized services through secondary databases held locally on CD ROM or magnetic tapes. The scope and functions of integrated library packages, till recently, were essentially restricted to providing access to documents at bibliographic level. The new versions of, integrated library packages, however, tend to provide additional features and functionalities akeen to digital libraries. Similarly, secondary information systems such as MEDLINE, INSPEC, COMPENDEX+ and CAS were essentially designed to serve as an effective tools for bibliographic control of research information. However, since these databases provide only bibliographic information on research articles, use had to depend heavily on physical collection available either in their institutional library or on inter-library loan from other libraries for references retrieved from the secondary services. Several attempts were made the in past to make the full-text of research articles available through online search services, although technology available till late 1980s and early 1990s supported only simple text (ASCII) without graphics. As such, more than 1,700 full-text journals were available through online search services like DIALOG and STN uptill 1989 although as simple text without graphics. Tools, techniques and protocols necessary for building-up digital libraries evolved with availability of computing power that allow parallel processing, multitasking, parallel consultation, parallel knowledge navigation and software tools that facilitate artificial intelligence and interactivity. Coincided with the availability of software, hardware and networking technology, the advent of the world wide web (WWW), its ever increasing usage and highly evolved browsers have paved the way for the creation of digital libraries. With rapid developments in the technologies necessary for developing digital libraries, the world of digital information resources has expanded quickly and exponentially. An Increasing number of commercial and society publishers are using the Internet as a global way to offer their publications to the international community of scientists and technologists. Resultantly, increasingly large numbers of STM (Science, Technology and Medical) electronic journals are appearing on the web. Digital information resources include not only rapidly growing collections of electronic full text resources, but also images, video, sound, and even object of virtual reality.

The most significant shift is in building digital collections is greater interoperability among information systems across the networks. With the technology available at an effordable cost, the libraries are initiating small digitization projects as individual library or as a group of libraries. Building-up digital collection and infrastructure required to access them is a challenge that every library has to deal with. Today's digital libraries are built around Internet and web technologies with electronic journals as their building blocks. The increasing popularity of Internet and developments in web technologies are catalyst to the concept of digital libraries. Figure 1 is a pictorial representation of digital

library infrastructure and services that can be generated from them.

Growth and developments in digital libraries can generally be attributed to emergence of Internet, particularly world wide web (WWW) as a media of information delivery and access, availability of highly evolved, extraordinarily simple and intuitive user interface, i.e. Internet Explorer and Chrome and advances in online storage technologies enabling storage of large amounts of contents at increasingly affordable cost. The product and services that goes into digital library comes from electronic publishing.

Several terms have been coined at different times to represent the concept of library without books, libraries having information in computer-readable format or having access to information in digitized or digital format. The terms which have been in vogue at different times include: paperless library, an electronic library, virtual library, library without boundaries and more recently digital library. The term digital library, at one hand, is used to refer to a system or applications whose function is chiefly to extend electronic access to material available in a conventional library to remote user, on the other hand, it is used to describe both commercial and academic systems designed to enable electronic access to a large corpora of electronic documents to authorized users.

The term digital library may mean different things to different people. It has been applied to an extraordinary range of applications and is frequently used to denote one or more of the followings:

- Collections in which complete contents of documents (as opposed to bibliographic citation or abstracts) are created or converted to computer processible form for online access;

- Providing digital access to material that already exists within traditional library collections, i.e. libraries of scanned image, images of photographic or printed texts, digital video segments;

- Scientific data sets like protein sequences or nucleic acid sequences, etc. Software libraries or multimedia works are often referred to as a digital library;

- Online databases and CD ROM information products, particularly those with multimedia or interactive video components or those which contain the complete contents of books or other publications;

- Computer storage devices on which information repositories reside, such as optical discs, juke boxes, CD ROM / DVD ROM towers, etc.;

- Database, including library catalogue accessible through the Internet; and

- Digital audio, video clips or full-length movies.

The only thing common about the range of products and services mentioned above are their being “digital” or “digitized”. While some of the above mentioned products and services qualify to be digital libraries, others do not qualify the characteristics and definition of a digital library given later in this chapter. The relatively recent use of the

term “digital library” can be traced to the “Digital Library Initiatives” funded by the US National Science Foundation, the Advanced Research Projects Agency, and the National Aeronautics and Space Administration (NASA) in the United States. In 1994, these agencies granted US$ 24.4 million to six universities in US for digital library research impelled by the sudden explosive growth on the Internet and web technology. The term was quickly adopted by the computer scientists, librarians and others. Thus, while the term “digital library” is relatively new, the concept behind the term and information resources consisting of digitized resources has a history spanning several years.

A digital library is not merely a collection of electronic information, it is an organized system of information that can serve as a rich resource for its user community. The library and information science community treat digital libraries as “logical extension and augmentation of physical libraries in the electronic information society (Marchionin 1998). The digital library extend and augment their physical counterparts by extending existing resources and services and enable development of new possibilities for information access and retrieval (Fox, 1998).

2. Traditional Library, Information Retrieval System and Digital Library

The services and collection in a traditional library are built around its physical possessions consisting of books, journals, microforms, video and audio cassettes, technical reports, theses & dissertations, standards and patents, etc. The primary purpose of a Library OPAC is to indicate the physical location of a document in the library. In a traditional library environment consisting of physical collections, it is necessary for a user to either come to the library or get the document in order to use it. Moreover, only one person at a time can use a physical document. However, traditional libraries offer additional social and educational benefits. Besides, most traditional libraries also offer hybrid services. Digital library removes physical restrictions that prevail in traditional libraries, provides multiple access, multiple listings and electronic transmission of its collection. Moreover, newer Web 2.0 / Library 2.0 applications now enable digital libraries to offer possibilities of “social networking” and “tagging” in web environment thereby imitating some of the social and educational benefits offered by traditional libraries. Digital libraries, however, come with complications such as: intellectual property, rights management, digital preservation, licenses and terms and conditions, etc.

Information retrieval systems (IRS) can be considered as precursor to the digital libraries. IRS is built with bibliographic databases as target for searching and retrieving textual information stored in them. The process of searching an information retrieval system is based on exact matching from string of text stored in bibliographic database using Boolean and proximity operators. Mistakes in the IRS system at the time of data entry or in search query results in mismatch. Digital libraries, in contrast, are based on pattern searching and inexact matching. While IRS provides metadata access only, digital library provides access to metadata and data. Migration from information retrieval system to digital libraries coincides with the development of full-text e-resources and spread of World Wide Web (WWW).

3. Electronic, Virtual, Hybrid and Digital Libraries

While the terms digital libraries and electronic libraries are used interchangeably and synonymously, the term "virtual library" or "library without wall" usually refers to the meta resources or subject portals that extend virtual accessibility of digital collections from several diverse sources without the users even knowing where the resource actually resides. Unlike digital libraries, virtual libraries do not consist of full-text resources, instead they are more like an index of relevant, hand-picked links to resources available on the Web. A virtual library could potentially be enormous, linking huge collections from all around the globe, or it could be very small, consisting of a few hundred links to digital resources maintained by an individual. The concept of "Hybrid Library" (Rusbridge, 1998) reflects the realities being faced by libraries as they attempt to integrate electronic resources acquired on CD ROM or other media or electronic access that they buy with the digital collections produced in-house. The hybrid library can be considered as a transitional phase between the conventional and digital library, where electronic and paper-based information sources are used alongside each other. The challenges associated with the management of hybrid library is to encourage end-user resource discovery and information use, in a variety of formats and from a number of local and remote sources in a seamlessly integrated way (Schawrtz, 2000). The hybrid library should be designed to bring range of technologies from different sources together in the context of a working library. In effect, a hybrid library maintains all or a major part of its collections in computer-processible form as an alternate or to supplement or to complement the conventional printed materials that exist in the libraries. It has a web-enabled computerized catalogue (WebPAC) accessible through the Internet and most of other in-house services like acquisition, books processing, circulation are computerized. A hybrid library has a strong presence on the Internet with a Home Page for the Library providing an integrated access interface, not only to digital collections available locally, but also to other commercial and non-commercial web-based digitized collections accessible to the library across the world.

4. Characteristics of Digital Libraries

A digital library promises a one-step, equitable and timely access to vast amount of diverse resources in a shared mode in a given specialty lifting traditional barriers of time and space. Digital libraries have the following characteristics associated with them:

- Digital libraries are the digital counterparts of traditional libraries and include both electronic (digital) as well as print and other (e.g. audio, video, graphics, animation, etc.) material;

- Digital libraries are not bound to physical spaces. Different components of digital library may be distributed to different locations that works coherently so as to meet the requirement of users;

- Requirement of physical spaces in digital environment reduces essentially for i) housing servers for hosting digital content; ii) PCs as clients for accessing digital content; and iii) staff for maintaining digital libraries.

- A digital library owns and controls the information, it provides access to information, not just a pointer to it;

- A digital library has a unified organizational structure with consistent points for accessing the data;

- A digital library is not a single entity, it may also provide access to digital material and resources from outside the actual confines of any one digital library;

- Digital libraries support quick and efficient access to a large number of distributed but interlinked information sources that are seamlessly integrated;

- Digital libraries offer access to its content to multiple users simultaneously, these content can be listed in multiple ways by different users simultaneously;

- Digital libraries have collections that i) are large and persist over time; ii) are well-organized and managed; iii) contain many formats; iv) contain objects and not just their representations; v) contain objects that may be otherwise unobtainable; and vi) contain some objects that are born digital; and

- Digital libraries include all the processes and services offered by traditional libraries though these processes will have to be revised to the accommodate difference between digital and paper media.

5. World Wide Web (WWW) V/s Digital Library

The World Wide Web (WWW) or the Web is a collection of thousands and thousands of documents and is considered as a digital library by many people. The web is means by which most digital libraries are accessed, but it is not a digital library itself although it has several features of a digital library. The web, unlike a digital library, is an unorganized collection of documents, many of them ephemeral information which does not have any durability or lasting value. Most search engines hunt down their holdings from web sites distributed across the web space, whereas digital libraries are generally more tightly controlled, and have a targeted customer set.

Today's digital libraries are built around Internet and web technologies. While the Internet serves as the carrier and provides the contents delivery mechanism, the web provides the tools and techniques for content publishing, hosting and accessing. The increasing popularity of Internet and developments in web technologies are catalyst to the concept of digital libraries. Further, availability of computing power that allow parallel processing, multitasking, parallel consultation and parallel knowledge navigation, put together, creates a semblance of artificial intelligence and interactively necessary for developing a digital library. Coincided with the availability of software, hardware and networking technology, the advent of World Wide Web (WWW), its ever increasing usage and highly evolved browsers have paved the way for creation of a global digital library.

6. Digital Library: Towards a Definition

The Association of Research Libraries, (Waters, 1998) one of the leaders in collaborative digitization programs in US, assigns following tenets to a digital library:

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

Borgman (1992) emphasized that digital libraries should not be viewed only as a point of access to digital information, but as a combination of

- a services;
- an architecture;
- a set of information resource, databases of text, numbers, graphics, sound, music or animation, etc.; and
- a set of tools and capabilities to locate, retrieve and utilize the information resources available.

7. Why Digital Library?

The unprecedented surge of activities and interest in digital library can generally be attributed to the following three factors:

i. Emergence of Internet and web technologies as a media of information delivery and access. The Internet, particularly world wide web (WWW), allows rapid access to a wide variety of networked information resources extending a uniform interface to a vast number of multimedia resources. The web, being a hypermedia based system, allow linking amongst electronic resources;

ii. Availability of highly evolved, extraordinarily simple and intuitive user interface, i.e. Internet Explorer and Netscape Navigator for all prevalent platforms; and

iii. Advances in online storage technologies enabling storage of large amounts of contents at increasingly affordable cost.

The digital library offers significant and unparallel improvement and value addition to library services while providing workable solutions to problems traditionally associated with the management of print-based collections in traditional libraries. Improved information retrieval and enhanced document delivery capabilities are widely acclaimed strength of digital libraries. Moreover, the cost of creating, storing, manipulating and transmitting digital information has decreased considerably providing necessary impetus to the digital library initiatives world wide. Rising acquisition and subscription fees have forced the libraries to find other means to make information available to their users and content aggregators and electronic publishers are providing means to do so.

Several large-scale digitization projects are aimed at conserving and preserving old, fragile and deteriorating documents of high scholarly value not only for preserving them but also for providing increased access and search possibilities that become possible once the documents are available in computer-processible form. Digital libraries enable greater access to digital contents, can be managed from remote locations and provide a way to enrich the teaching and learning environment. Since information in digital library is electronically stored and accessed, it is not bound to space and time. Digital library systems can be accessed simultaneously by multiple users guaranteeing continuous availability of documents. Digital library implementation can dramatically reduce floor space requirements as compare to conventional shelf-type storage of books and journals

8 Some Important Digital Libraries

New Zealand Digital Library (http://www.nzld.org/)

The New Zealand Digital Library, maintained by the University of Waikato, provides web access to several document collections, with varied subject content, languages and formats. It includes historical documents, humanitarian and development information, computer science technical reports and bibliographies, literary works, and magazines. Content formats include text (ASCII, PostScript, PDF), graphics, audio and video. The NZDL supports a simple but powerful bibliographic and full-text search and browse interface, including hierarchical browsing and display. NZDL has given a lot of importance to structuring the search, browse and display interfaces, making them user friendly. Metadata (author, title, keywords, etc.) plays a key role in supporting field-based searches and browsing. The NZDL has been built using the Greenstone Digital Library software, developed out of a research programme at the University of Waikato. The Greenstone Digital Library software (GSDL) is available freely as open-source, under the terms of GNU public license

Networked Computer Science Technical Reference Library (http://www.ncstrl.org/)

The Networked Computer Science Technical Reference Library (NCSTRL) is an international collection of computer science research reports and papers made available by more than 100 participating institutions worldwide. The Majority of the institutions are universities and research laboratories. While the full-text reports are maintained on servers in participating institutions, a central index of bibliographic details is maintained for all full-text records with searching and linking facilities. The index is updated automatically when a new report is added to any of the institutional servers. New institutions can join NCSTRL and use free software made available freely. The NCSTRL is being shifted to OAI-compliant repository using E-prints software for institutional repositories and ARC for harvesting metadata from distributed repositories.

transmitting digital information has decreased considerably providing necessary impetus to the digital library initiatives world wide. Rising acquisition and subscription fees have forced the libraries to find other means to make information available to their users and content aggregators and electronic publishers are providing means to do so.

Several large-scale digitization projects are aimed at conserving and preserving old, fragile and deteriorating documents of high scholarly value not only for preserving them but also for providing increased access and search possibilities that become possible once the documents are available in computer-processible form. Digital libraries enable greater access to digital contents, can be managed from remote locations and provide a way to enrich the teaching and learning environment. Since information in digital library is electronically stored and accessed, it is not bound to space and time. Digital library systems can be accessed simultaneously by multiple users guaranteeing continuous availability of documents. Digital library implementation can dramatically reduce floor space requirements as compare to conventional shelf-type storage of books and journals.

- These linkages between many digital libraries and information services are transparent to end-users;

- Universal access to the digital libraries and information services is the goal; and

- Digital Library collections are not limited to documents surrogates, they also include digital artifacts that cannot be represented or distributed in printed formats.

Borgman (1992) emphasized that digital libraries should not be viewed only as a point of access to digital information, but as a combination of

- a services;
- an architecture;
- a set of information resource, databases of text, numbers, graphics, sound, music or animation, etc.; and
- a set of tools and capabilities to locate, retrieve and utilize the information resources available.

Terence R. Smith (1997), defined digital libraries as "controlled collections of information bearing objects (IBOs) that are in digital form and that may be organized, accessed, evaluated and used by means of heterogeneous and extensible set of distributed services that are supported by digital technology".

Clifford Lynch (1995), a well-know expert on digital libraries and new technologies, defined digital library as "a system providing a community of users with coherent access to a large, organized repository of digital information and knowledge. The digital library is not just one entity, but multiple sources seamlessly integrated." Michael Lesk, who predicts that half of the materials accessed in major libraries will be digital by the early 21st century (Lesk, 1997), defines digital libraries as "organized collections of digital information that combine the structuring and gathering of information, which libraries and archives have always done, with the digital representation that computers have made possible. Digital information can be accessed rapidly around the world, copies for preservation without error, stored compactly, and searched very quickly. A true digital library also provides the principles governing whatis included and how the collection is organized" (Lesk, 1997).
Emphasizing management aspects of digital collections and services, Arms (2000) defines digital libraries as "managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network". Laying emphasis on digital technology, Oppenheim and Smithson (1999) define digital library as "an information service in which all the information resources are available in computer processible form and functions of acquisitions, storage, preservation, retrieval, access and display are carried out through the use of digital technologies.

Painting a multi-dimensional picture, Marchionini and Fox (1999) identified the following four dimensions of digital libraries:

i. **Community**: Reflects social, political, legal and cultural issues;

ii. **Technology**: includes technical progress in computing, networking, information storage and retrieval, multimedia, interface design, etc.;

iii. **Services**: includes present and future services, personalization, digital reference services, real-time question answering, on-demand help, information literacy and user involvement mechanisms; and

iv. **Content**: represents all possible kinds of forms and genre of information, printed as well as digital.

It is critical that digital libraries provide an organized and structured access to information contents in a distributed environment and assist users in searching, evaluating and utilizing resources irrespective of their format. Digital libraries combine collection and expertise in a seamless interface, and therefore, require specialized staff to select, organize, evaluate, interpret, offer intellectual access, preserve the integrity and ensure the persistence over time of digital works so that they are readily and economically available for use by a defined community or set of communities (Waters, 1992).

7. Why Digital Library?

The unprecedented surge of activities and interest in digital library can generally be attributed to the following three factors:

i. Emergence of Internet and web technologies as a media of information delivery and access. The Internet, particularly world wide web (WWW), allows rapid access to a wide variety of networked information resources extending a uniform interface to a vast number of multimedia resources. The web, being a hypermedia based system, allow linking amongst electronic resources;

ii. Availability of highly evolved, extraordinarily simple and intuitive user interface, i.e. Internet Explorer and Netscape Navigator for all prevalent platforms; and

iii. Advances in online storage technologies enabling storage of large amounts of contents at increasingly affordable cost.

The digital library offers significant and unparallel improvement and value addition to library services while providing workable solutions to problems traditionally associated with the management of print-based collections in traditional libraries. Improved information retrieval and enhanced document delivery capabilities are widely acclaimed strength of digital libraries. Moreover, the cost of creating, storing, manipulating andtransmitting digital information has decreased considerably providing necessary impetus to the digital library initiatives world wide. Rising acquisition and subscription fees have forced the libraries to find other means to make information available to their users and content aggregators and electronic publishers are providing means to do so.
Several large-scale digitization projects are aimed at conserving and preserving old, fragile and deteriorating documents of high scholarly value not only for preserving them but also for providing increased access and search possibilities that become possible once the documents are available in computer-processible form. Digital libraries enable greater access to digital contents, can be managed from remote locations and provide a way to enrich the teaching and learning environment. Since information in digital library is electronically stored and accessed, it is not bound to space and time. Digital library systems can be accessed simultaneously by multiple users guaranteeing continuous availability of documents. Digital library implementation can dramatically reduce floor space requirements as compare to conventional shelf-type storage of books and journals.

8. Some Important Digital Libraries

New Zealand Digital Library (http://www.nzld.org/)

The New Zealand Digital Library, maintained by the University of Waikato, provides web access to several document collections, with varied subject content, languages and formats. It includes historical documents, humanitarian and development information, computer science technical reports and bibliographies, literary works, and magazines. Content formats include text (ASCII, PostScript, PDF), graphics, audio and video. The NZDL supports a simple but powerful bibliographic and full-text search and browse interface, including hierarchical browsing and display. NZDL has given a lot of importance to structuring the search, browse and display interfaces, making them user friendly. Metadata (author, title, keywords, etc.) plays a key role in supporting field-based searches and browsing. The NZDL has been built using the Greenstone Digital Library software, developed out of a research programme at the University of Waikato. The Greenstone Digital Library software (GSDL) is available freely as open-source, under the terms of GNU public license.

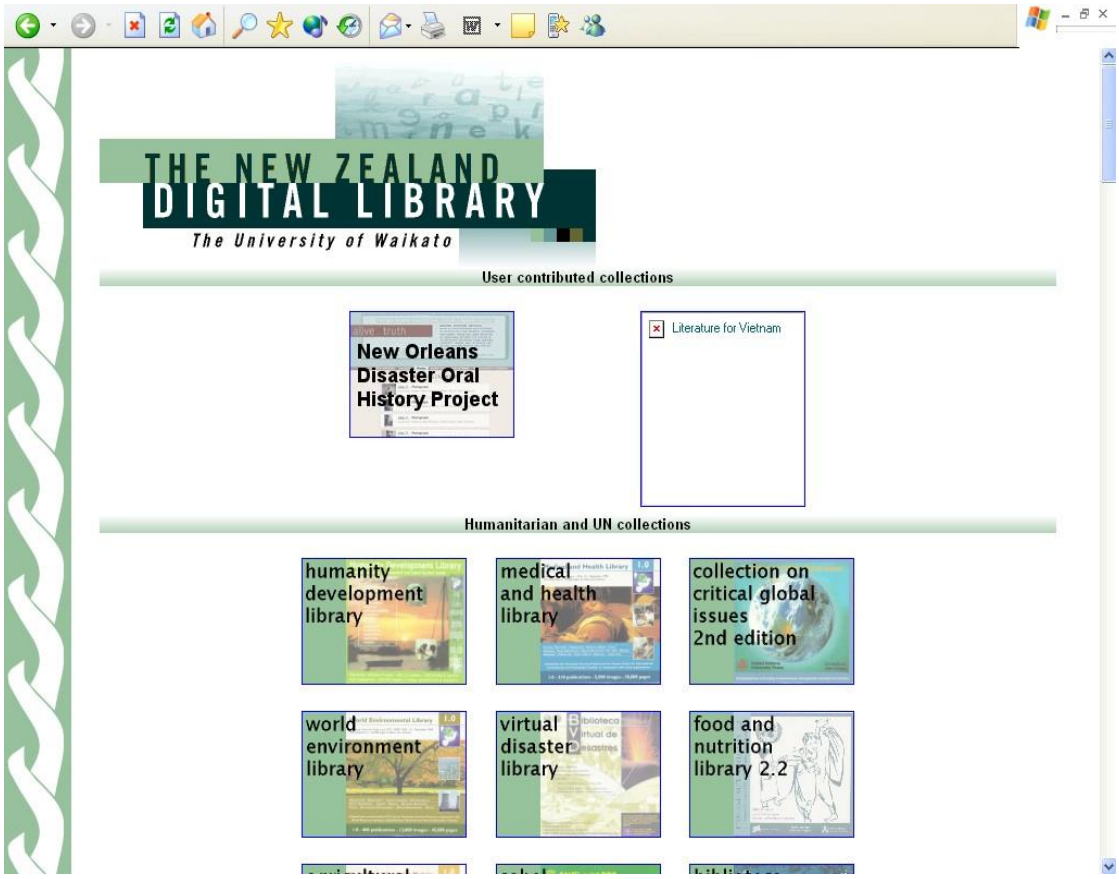


Fig. 1: New Zealand Digital Library (http://www.nzld.org/)

Networked Computer Science Technical Reference Library (http://www.ncstrl.org/)

The Networked Computer Science Technical Reference Library (NCSTRL) is an international collection of computer science research reports and papers made available by more than 100 participating institutions worldwide. The Majority of the institutions are universities and research laboratories. While the full-text reports are maintained on servers in participating institutions, a central index of bibliographic details is maintained for all full-text records with searching and linking facilities. The index is updated automatically when a new report is added to any of the institutional servers. New institutions can join NCSTRL and use free software made available freely. The NCSTRL is being shifted to OAI-compliant repository using E-prints software for institutional repositories and ARC for harvesting metadata from distributed repositories.

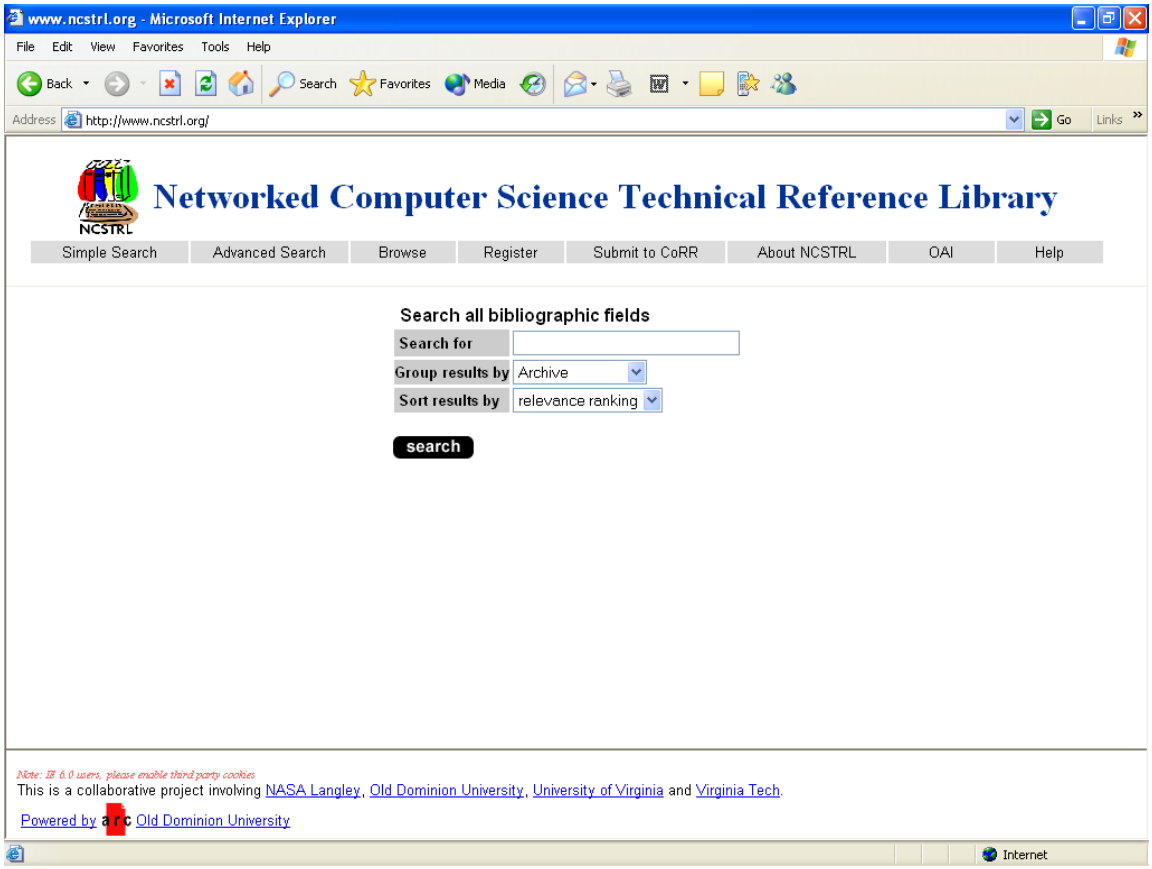


Fig. 2: Networked Computer Science Technical Reference Library (http://www.ncstrl.org/)

ArXiv.org (http://www.arxiv.org/)

The ArXiv (http://www.arXiv.org/), started in 1991 by Paul Ginsparg at Los Alamos National Laboratory, is the oldest eprint archive. The repository, now hosted at the Cornell University, has become a fundamental means of communication for a growing number of fields, starting with theoretical high-energy physics, later spreading to other areas of physics, and now also to computer science and mathematics. ArXiv is leading example of successful implementation of developments in information technology, which led to an alternative model of scholarly communication. This archive processes 35,000 submissions every year. It receives two-thirds of its two million weekly hits from institutions outside the United States, including many research facilities in developing regions. The arXiv has become indispensable to researchers world wide, but in particular to research institutions in developing countries. The success and widespread adoption of arXiv has prompted establishment of institutional archives and subject-based digital repositories in different disciplines. Scientists and librarians have become aware of benefits of open access archiving that is being considered as an alternative method of scholarly publishing. The Institute of Mathematical Sciences, Chenai, maintains the Indian mirror site of ArXiv.

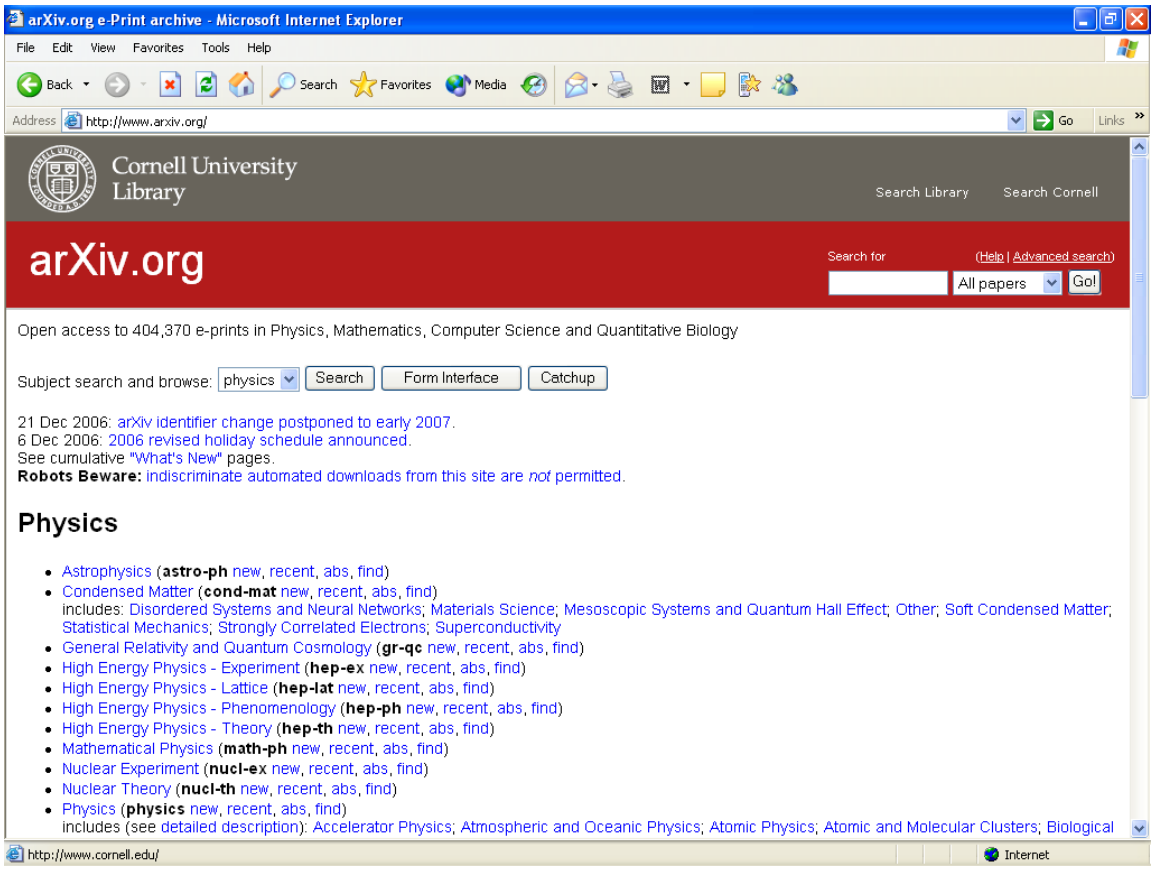


Fig. 3: ArXiv.org (http://www.arxiv.org/)

ScienceDirect (http://www.sciencedirect.com/)

ScienceDirect is the web-based interface to the full-text database of a commercial publisher, namely Elsevier Science, one of the world's largest providers of scientific, technical and medical (STM) literature. ScienceDirect contains over 25% of the world's science, technology and medicine information. It offers a rich electronic environment for research journals, bibliographic databases and reference works. The database offers more than 2,000 scientific, technical and medical peer-reviewed journals, over 59 million abstracts, over 7 million full-text scientific journal articles, an expanding suite of bibliographic databases and linking to another one million full-text articles via CrossRef to other publishers' platforms. In addition, the Backfiles program of ScienceDirect offers the ability to search a historical archive of over 6.75 million articles directly from the desktop of a user, back to Volume 1, Issue 1.

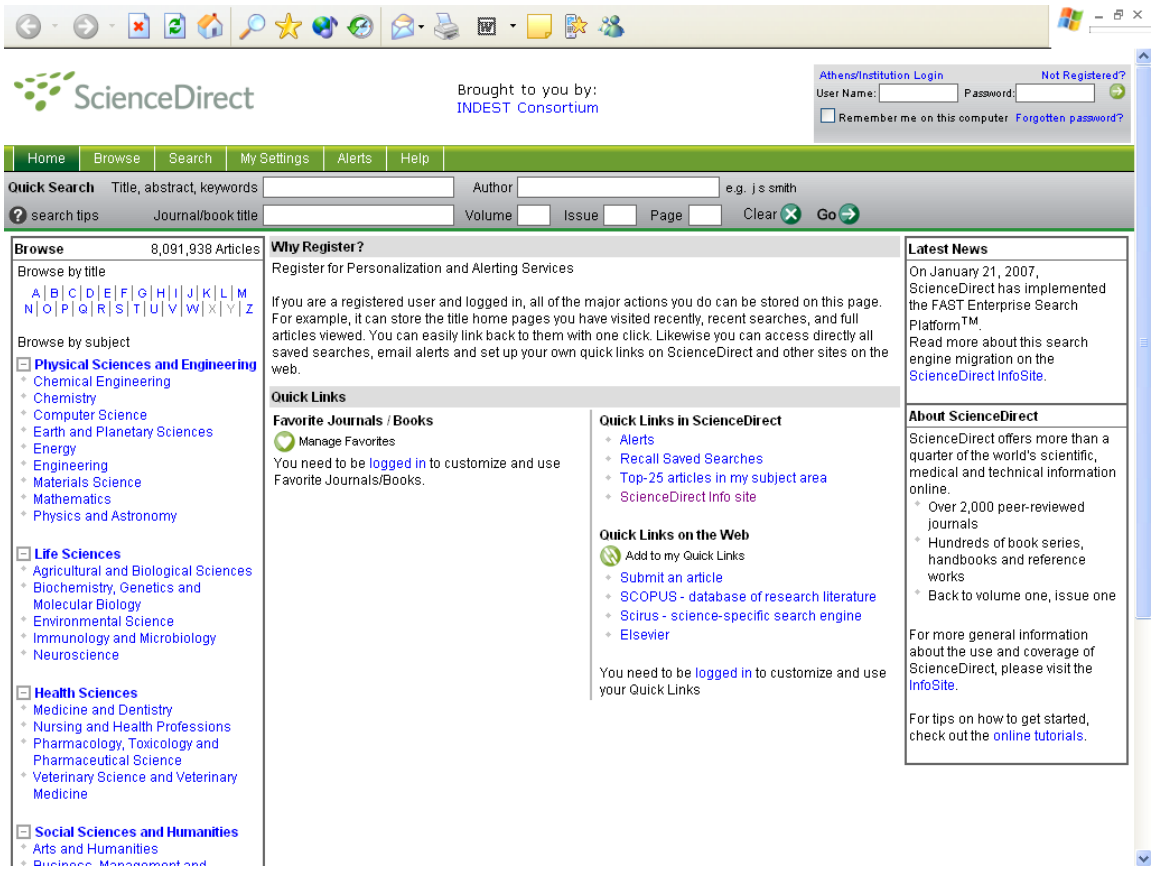


Fig. 4: ScienceDirect (http://www.sciencedirect.com/)

9. Summary

Digital libraries are amongst the most complex and advanced form of information systems. Deployment of digital library requires integration of several information technologies because of many diverse requirements involving creation of digital contents, its organization, ontology, development of interactive interfaces for users, multiple accesses and listings, digital document imaging, OCR, distributed database management, web technology, hypertext, information storage and retrieval system, experts system, intellectual property rights, integration of multimedia information services, management of multilingual collection, data mining, electronic and real-time reference service, electronic document delivery and personalization. Due to these unique challenges and opportunities, the digital libraries are emerging as a growing interdisciplinary area of research and education for information science, computer science, library science and a number of other related disciplines.

Various terminologies associated with development and evolution of hypertext, imaging technology, World Wide Web and other related technology are discussed. Basic characteristics of digital libraries are enunciated along with their definitions by some of the leading technologists in this field. The section deliberates upon the concept of “hybrid

library" which reflects the realities being faced by libraries as they attempt to integrate electronic resources acquired on CD ROM or other media or electronic access that they buy with the digital collections produced in-house. The hybrid library is in continuum between the conventional and digital library, where electronic and paper-based information sources are used alongside each other. It discusses the need for digital libraries and compares it with Web, traditional libraries and traditional information storage and retrieval system. Examples of different types of digital libraries conclude this section on the digital library. A glossary of terms used in the text is given to provide better understanding to the concept of digital libraries.

References

Arms, William Y. Digital libraries. Massachusetts, MIT Press, 2000.

Association of Research Libraries. ARL Proceedings 126: Annual Meeting, 17-19, 1995. (***http://arl.cni.org/arl/proceedings/126/2-defn.html***)

Fox, E. and Marchionini, Gary. Towards a worldwide digital library. Communication of the ACM, 41(4), 29-32, 1998.

Lesk, M. Practical digital libraries: Books, bytes and bucks. San Francisco, Morgan Kaufman, 1997.

Lynch, Clifford and Hector Garcia-Molina. IITA Digital Library Workshop, Reston, VA, May 18-19, 1995.

Marchionini, G., Plaisant, C. and Komlodi, A. Interfaces and tools for the Library of Congress National Digital Library Program. Information Processing and Management, 34(5), 535-555,1998.

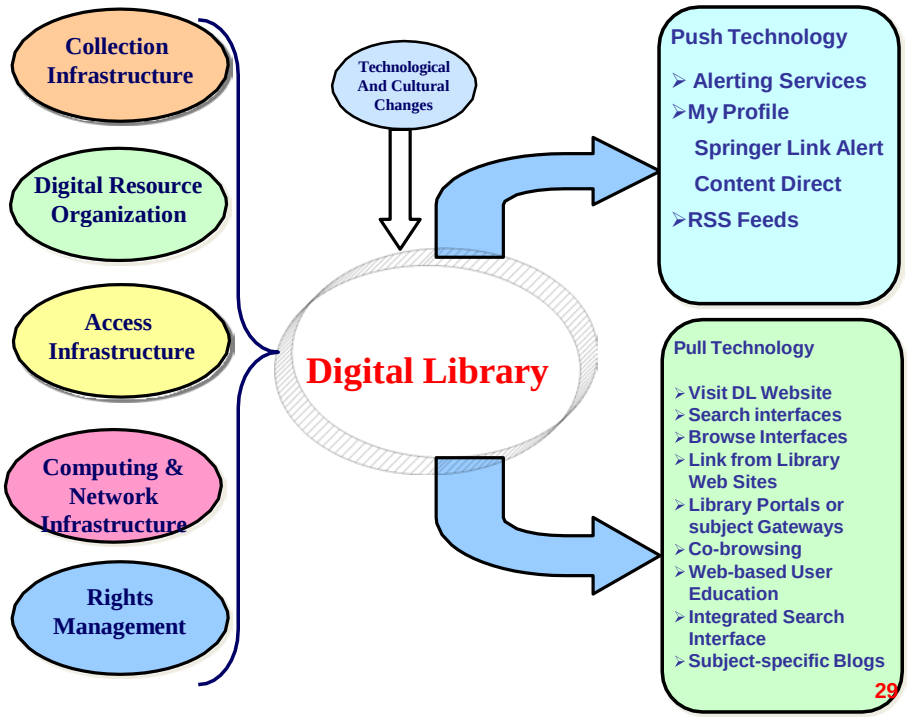
Oppenheim, C. and Smithson, D. What is the hybrid library? Journal of Information Science, 25(2), 97-112, 1999.

Rusbridge, Chris. Towards the hybrid library. D-Lib Magazine, July / August, 1998. (http://www.dlib.org/dlib/july98/rusbridge.html)

Schwartz, C. Digital libraries: An overview. Journal of Academic Librarianship, 26(6), 385-396, 2000.

Smith, T.R. Meta information in digital libraries. Int.J.Digital Libraries,1,105-107,1997.

Waters, D. Electronic technologies and preservation. European Research Libraries Cooperation, 2(3), 285-293, 1992.



Paper 8: Digital Libraries

Unit-2
Historical Evolution of Digital Libraries

1. Introduction

Although the term digital library has gained popularity in recent years, they have evolved along the technological ladder for the past thirty years. In the early 1970s, digital libraries were built around mini and main-frame computers providing remote access and online search and retrieval services to online databases using computer and communication technology available at that time.

The historical evolution of digital libraries has not been linear. Digital library is an eclectic science. Thus, contributions for evolution of digital libraries have come from several disciplines, leading to multiple conceptions of digital libraries,each one influenced by the perspective of primary discipline. History of digital libraries, therefore, is the history of variety of different types of information system and technologies that have been considered as "digital libraries" or their precursors. These systems and technologies are very heterogeneous in their objectives, scope and functionalities. As such, their evolution does not follow a single path. It can be observed that most of the systems described in this module are still practiced in one form or another using newer technological solutions and have applications in diverse field of information management.

The Rest of this module delves upon the historical evolution of different systems and technologies that contributed to the evolution of digital libraries.

2. Computers and Microprocessor Technology

Development in digital library progressed along with the developments in computer and communication technology. Two significant computers were built in 1946 and 1947. The ENIAC I (Electronic Numerical Integrator and Calculator) computer was developed by John Mauchly and J. Presper Eckert at the University of Pennsylvania. It contained over 18,000 vacuum tubes, weighed thirty tons and was housed in two stories of a building. Another computer, EDVAC, was designed to store two programs at once and switch between the sets of instructions. A major breakthrough occurred in 1947 when Bell Laboratories replaced vacuum tubes with the invention of the transistor. The transistors decreased the size of the computer, and at the same time increased the speed and capacity. The UNIVAC I (Universal Automatic Computer) became the first computer using transistors and was used at the U.S. Bureau of the Census from 1951 until 1963. Software development also was in progress during this time. Operating systems and programming languages were developed for the

computers being built. Invention of integrated circuit by Robert Noyce of Intel and Jack Kirby of Texas Instruments in 1960s. All the components of an electronic circuit were placed onto a single "chip" of silicon(Bosworth, 2011).

Dramatic reductions in the size and cost of computer components and equally impressive gains in the speed, storage capacity and reliability of hardware components have expanded their use rapidly in all activities and functions of a library and information centre. Notable reductions in the size of microprocessors combined with dramatically enhanced capacity have added new dimensions to the computer hardware technology. Initially, small silicon chips contained only few components and circuits, but the average number of components has doubled each year since 1965. Early small-scale integration efforts first gave way to large-scale integration (LSI) chips that contained thousands of components, very large-scale integration (VLSI) chips that contained hundreds of thousands of components and circuits, and now ultra large-scale integration (ULSI) chips come with millions of components and circuits.

A microprocessor, also known as Central Processing Unit (CPU) of a computer, is a complete computation engine that is fabricated on a single chip. Different companies like Intel, Advanced Micro Devices (AMD), Motorola etc. manufacture microprocessors. However, Intel manufactures most widely used microprocessors. The first microprocessor used in a PC was Intel 8080. Introduced in 1974, it was a complete 8-bit microprocessor on a single chip. Intel 8088 used in the IBM PC in 1979, was the first microprocessor that made its presence felt in the market. The PC market moved from the 8088, 80286, 80386, 80486, Pentium, Pentium-II, III, IV and now to Intel Dual Core and Quad Core.

3 Digital Storage Technology

Like microprocessor technology, digital storage devices have also witnessed notable reductions in its size and cost with dramatically enhancement in capacity of storage. There are two types of data storage devices, i.e. removable data storage devices and non-removable data storage devices. The data storage devices come in many sizes and shapes. The storage devices can also be categorized based on media used for storage, for example magnetic storage media, optical storage media, metal-oxide semiconductors or flash memory devices (popularly known as Pen drives or USB drives).

Magnetic Storage Media

Magnetic storage media are commonly used for large volumes of data (e.g., video, image, or remote sensing data). The first public demonstration of a magnetic audio recorder was invention of Valdemar Poulsen in 1898(Wikipedia, 2014). While the early magnetic storage devices were designed to record analog audio signals, most audio and video magnetic storage devices including computer store digital data. Large amounts of data are stored through tape drives because the capacity on the drives is huge - three billion (or three gigabits) of data per square inch can fit on a single magnetic disk. Hard discs, floppies, tapes, cartridges, etc. are example of magnetic media.

Optical Storage Media

Although research into optical data storage is going on for several decades, the first popular optical storage media, called the compact disc, was introduced in 1982. With release of Yellow Book in 1985, two versions of CD ROM were released in 1988, i.e.
i) CD-R: write-once, read-many (data once written cannot be erased); and ii) DR-RW: the data once written can be erased completely and the same storage device can be used again for storing different data. A typical disc used in a computer-based CD drive stores 700 MB. (Wikipedia, 2014)

DVD (initially "Digital Video Disc", then modified to "Digital Versatile Disc"), like a CD, is an optical storage system was rolled out in 1996 as successor to CD. Recordable formats of DVD called DVD-R,was released in 1997 and writable format, called DVD+R,was released in 2002(Wikipedia, 2014). The DVD format provides several configurations of data layers, moving from 2D storage to 3D storage. Each configuration is designed to provide additional storage capacity. Commonly used DVDs (one side- one layer) in contrast to CDs has storage capacity of 4.7 GB.

Flash Memory Devices or USB Drives

Invention of metal-oxide semiconductors-based storage devices, popularly known as pen drives or USB drives or flash memory devices were patented in 1999 by IBM(Wikipedia, 2014). USB drive is a non-volatile computer memory that can be electrically erased and reprogrammed. It is a technology that is primarily used in memory cards and USB flash drives for general storage and transfer of data between computers and other digital products. Flash memory stores information in an array of memory cells made from floating-gate transistors. In traditional single-level cell (SLC) devices, each cell stores only one bit of information. Some newer flash memory, known as multi-level cell (MLC) devices, can store more than one bit per cell by choosing between multiple levels of electrical charge to apply to the floating gates of its cells. Kingston Technology Company has released 3.0 USB Flash drives in 2013 with 1TB storage capacity(Kingston Technology Company, 2014).

4 Online Databases and Information Retrieval System (IRS)

Creation and remote accessibility of online databases through Information retrieval systems (IRS) can be considered as an important landmark and precursor to the digital libraries. The first databases were bibliographic in nature and were online version of existing indexing and abstracting services such as Biological Abstracts, Index Medicus, Chemical Abstracts, etc. These databases used data files generated in the process of electronic phototypesetting of printed abstracting and indexing services and other primary journals. These bibliographic databases were hosted on mini and main-frame computers providing remote access and online search and retrieval services to users using computer and communication technology available at that time (Arora, 1996).

The earliest application of digital library concepts involved character-coded storage and full-text indexing of legal and scientific documents. The Legal Information through Electronics (LITE) System was first implemented by the US Air Force in 1967. DIALOG became the first commercial online service in 1972. By 1975 there were 300 publicly available online databases. By 1988, 3,893 online online databases were available from 1,723 database producers and 576 online services.

Sophisticated information storage and retrieval systems were built during 1980s using state-of-the-art technology of distributed database management system linking different remote systems. By the year 1988, only half of all databases were bibliographic. With the introduction of a number of online databases containing textual information, news, statistics, commodity prices, etc., a third type of databases holding text of full-length documents started appearing. As such, online hosts like DIALOG, STN, BLAISE and ESA-IRS were not only offering online databases but also full-text online journals for past several years, although as a simple ASCII or text files without graphics and pictures. In1989, there were almost 1,700 full-text sources in sixteen online systems. (Arora, 1996)

Availability of CD ROM in late 1980s, as a media with high storage capacity, longitivity, and ease of transportation triggered production of several CD ROM information products which were earlier

available through online vendors or as conventional abstracting and indexing services in printed format. Moreover, several full-text databases also started appearing in late 1980s and early 1990s, launching the beginning of digital era. Some of the important full-text digital collections available on CD ROM include: ADONIS, IEEE / IEE Electronic Library (IEL), ABI/INFO, UMI's Business Periodical On disc and General Periodicals On disc, E space World, US Patents, etc.

5 Computer-based Information Storage and Retrieval System

Several software packages were released during mid-1970s and late 1970s for computer-based storage, indexing and retrieval of documents in character-coded form. Some of the better known text storage and retrieval packages included: IBM's Storage and Information Retrieval System (STAIRS), Battelle Automated Search Information System (BASIS), INQUIRE, BRS/SEARCH, DOCU/MASTER, ASSASSIN,
STATUS, CAIRS, etc. By the late 1980s, text storage and retrieval programs were available from dozens of vendors for major computing environment including main- frame, microcomputers and LAN. Micro-CDS/ISIS, one of such advanced non- numerical information storage and retrieval software developed by UNESCO in 1985, was used extensively by libraries especially in developing countries. Micro CDS/ISIS is currently available in different flavours including CDS/ISIS for Windows, GenISIS, JavaISIS, WEBLIS, WWW-ISIS, etc.

Availability of a wide range of Database Management System (DBMS) such as Ingres, Microsoft Access, MS-SQL Server, MS FoxPro, MySQL, NoSQL, Oracle, Postgres, SqLite and MongoDB in late 1990s and early 2000 also contributed to evolution of digital libraries.

6 Digital Imaging Technology

Digital imaging was developed in 1960s and 1970s to avoid the operational weaknesses of film cameras for use in scientific and military missions. The first digital image was produced in 1920, however, the invention of the CCD (charge- coupled device) in 1969 at AT&T Bell Labs by Willard Boyle and George E. Smith led to its application of imaging technology in consumer products like digital scanners and digital camera(Wikipedia, 2014).

Digital document imaging system, which employ computer hardware and software to scan and store images of documents in digitized formats, were evolved in early 1980s to overcome the limitation of text storage and retrieval systems which could only store textual information. The earliest application of a document imaging system was the "Optical Disk Pilot Project" at the Library of Congress. Several document imaging software packages are currently available in the market.

7. Internet Technology and its Services

The history of Internet (Arora, 2004) can be traced back to 1957 when erstwhile Soviet Union launched its first satellite, Sputnik I, prompting US President Dwight Eisenhower to launch Defence Advanced Research Projects Agency (DARPA) to regain the lead in the technological race. DARPA's mission was to advance science and technology for military applications. The DARPA developed its first successful satellite in 18 months. By the end of 1960, it began to focus on computer networking and communication technology essentially to established communication links between research centres and universities across the country as part of its overall mission. ARPANET was

commissioned in 1969 and by 1971 it had 15 nodes and 23 hosts. The e-mail was invented in 1972 by Ray Tomlinson to send messages across a distributed network. In 1973, the first international connection to the evolving Internet was established at the University College of London and the Royal Radar Establishment (Norway). In the same year, DARPA initiated a research program to investigate techniques and technologies for interlinking packet networks of various kinds. The objective was to develop communication protocols based on "packet- switching" that would allow networked computers to communicate seamlessly across multiple, geographically dispersed locations. The "packet-switching" would split the data to be transmitted into tiny packets that can take different routes to their destination. This was called the Internetting project and the system of networks that emerged from the research was known as the "Internet". The system of protocols which was developed over the course of this research effort became known as the TCP/IP Protocol Suite, after the two initial protocols developed: Transmission Control Protocol (TCP) and Internet Protocol (IP).

The operation management of the emerging Internet was handed over to the Defence Communication Agency (DCA) in1975. Unix to Unix Copy Program (UUCP) was developed at the Bell Labs (AT & T) in the year 1976. 1977 witnessed the development of mail specifications (RFC 733). Usenet was established in the same year using UUCP (Unix to Unix Copy Program) between Duke and the University of North Carolina (UNC). DARPA also established the Internet Configuration Control Board (ICCB) in the year 1977.

8. Electronic Resources

The first electronic resources, in true sense, appeared in the form of bibliographic records in libraries. Machine Readable Cataloguing (MARC), introduced in 1964, can be considered as a major development in this regard. Soon after, automation of libraries started in a big way in the 1970s with the introduction of integrated library automation packages. The trend picked-up in the early 1980s with the introduction of PCs at a cost affordable to the libraries. The computerized catalogues of individual libraries led to formation of union catalogues through library networks like OCLC that were developed to facilitate online and copy cataloguing and resource sharing. By early 1970s, library OPACs and union databases were accessible from remote locations. Moreover, online search services, like DIALOG, ORBIT, BRS Search and Datastar in USA; BLAISE and Pergamon Infoline in UK; DIMDI in Germany; Euronet and Diane in Europe; ESA-IRS in Italy; and CAN/OLE in Canada, etc. were also made accessible online to the research community. Appearance of bibliographic and full-text databases on CD ROM by late 1980s can be considered as a major breakthrough in the evolution of electronic resources. Most of the bibliographic databases that were accessible through the online search services like DIALOG and STN became available on CD ROM (Arora, 2007).

The emergence of Internet and the World Wide Web (WWW) in early 1990s, as a new media of information storage and delivery, came as a real boon for evolution of electronic resources. While searching bibliographic databases became popular, it created demand for actual content in full-text that became difficult for libraries to obtain. Coincided with evolution of World Wide Web (WWW), display technology evolved, cost of storage came down drastically and networks became faster. It became possible for publishers to deliver content, either as a bitmap page images or other structured formats such as HTML, PDF or RTF. Increasingly larger number of publishers started using the Internet as a global way to offer their publications to the international community of scientists and technologists given the fact that technology is in a position to deliver more content to more users at a significantly lower cost per user. These new technologies are continuously driving the electronic resources to new peaks of usage, significantly beyond the library's subscribed content.

These Internet and web technologies brought in the graphical components in electronic resources and digital libraries that were missing earlier.

There has, thus, been a steady move up the technological scale for the electronic resources from early (late 1980s) low-end electronic publications available as ASCII files, to being organized and searchable on gophers (1992), and to being tagged and graphically viewable on World Wide Web sites (1994). In 1981, CSNET (Computer Science Network) was built with the collaboration of a number of universities and industries in USA. The National Science Foundation gave financial support to the CSNET to provide networking services. CSNET used the Phonenet MMDF protocol for telephone-based electronic mail relaying and, in addition, pioneered the first use of TCP/IP over X.25 using commercial public data networks. The CSNET server provided an early example of white pages directory service and this software is still in use at numerous sites. At its peak, CSNET had approximately 200 participating sites and international connections to approximately fifteen countries. Another important development in the same year was the creation of BITNET (Because it's time network). The BITNET was started as a cooperative network at the City University of New York with the first connection to University of Yale. At its peakin 1991, BITNET was connected to almost 500 organizations and 3,000 nodes in educational institutions in North America, Europe (as EARN), Israel (as ISRAEARN), India (TIFR) and some Persian Gulf states (as GulfNet). It was also very popular in other parts of the world, especially in South America, where about 200 nodes were implemented and were heavily used in the late 1980s and early 1990s. With the rapid growth of TCP/IP systems and the Internet in the early 1990s, and phasing out of IBM mainframe that was the base platform, BITNET's popularity and use diminished quickly. In 1996, CREN ended their support for BITNET. As of 2007, BITNET has essentially ceased operation.

1982 was a year of great significance in the growth and development of Internet. Defence Communication Agency (DCA) and DARPA adopted Transmission Control Protocol (TCP) and Internet Protocol (IP) suite (commonly known as TCP/IP) as the official protocol suite for ARPANET. This led to one of the first definition of Internet as connected set of networks using TCP/IP. In the same year, the Eunet (European UNIX Network) was created to provide e-mail and Usenet services in Europe. The External Gateway Protocol (EGP) was also developed in the same year, which defines protocols for connecting networks that were not based on TCP/IP with the Internet. The University of Wisconsin developed "Name Server" in 1982 that facilitated translation of names into strings of numbers. This development led to the practice of assigning domain names for the sites that is being practiced even now. Other significant development that took place in 1982 included splitting of ARPANET into ARPANET and MILNET. The MILNET was later integrated with the Defence Data Network created in 1981.

Launch of desktop computers in1982 led to major shift from having a single, large main frame computer connected to the Internet on each site to the entire local areas network connected to the Internet. In the same year, the Internet Activities Board (IAB) replaced ICCB with a primary mission to guide evolution of the TCP / IP protocol suite and to provide research advice to the Internet community.

Domain Name Servers as distributed databases were introduced in 1984 to facilitate translation from domain names to IP addresses. Transition to naming standards from numeric addresses proved to be very helpful in popularisation of the Internet. For example, it is much easier to remember www.yahoo.com than its numerical equivalent.

In 1986, the U.S. National Science Foundation (NSF) initiated the development of the NSFNET, which, today provides a major backbone communication service for the

Internet. The National Aeronautics and Space Administration (NASA) and the U.S. Department of Energy contributed additional backbone facilities in the form of the NSINET and ESNET respectively.

The Network News Transfer Protocol (NNTP) was designed to enhance news performance over TCP/IP.

In 1987, the NSF signed a cooperative agreement to manage the NSFNet backbone with Merit Networks, Inc. Merit, IBM and MCI later founded Advanced Network and Services, Inc. (ANS). In the same year, BITNET and CSNET merged to form the Corporation for Research and Educational Networking (CREN). In the fall of 1991, CSNET service was discontinued having fulfilled its important early role in the provision of academic networking service. A key feature of CREN is that its operational costs were fully met through dues paid by its member organizations.

A computer virus for the first time affected approximately 6,000 of total 60,000 hosts on the Internet in the year 1988. The vulnerability of Internet and the need for more security was realised for the first time. DARPA formed the Computer Emergency Response Team (CERT) in response. In the same year, the Department of Defence adopted Open Systems Interconnection (OSI).

The total number of hosts on the Internet rose to 100,000 in 1989. The year also witnessed first relays between a commercial electronic mail carrier and the Internet. MCI Mail connected through the Corporation for the National Research Initiative (CNRI) and CompuServe connected through Ohio State University. The Corporation for Research and Education Networking (CREN) was formed with the merger of CSNET and BITNET. The Internet Engineering Task Force (IETF) and Internet Research Task Force (IRTF) also came into existence under the IAB in the year 1989. In the same year, several other countries got connected to the NSFNet including Australia, Germany, Israel, Italy, Japan, Mexico, the Netherlands, New Zealand, Puerto Rico and the United Kingdom. In Europe, major international backbones such as NORDUNET and others provide connectivity to over one hundred thousand computers on a large number of networks. During the course of its evolution, particularly after 1989, the Internet system began to integrate support for other protocol suites into its basic networking architecture. The present emphasis in the system is on multi-protocol internetworking, and in particular, with the integration of the Open Systems Interconnection (OSI) protocols into the architecture.

During the early 1990's, OSI protocol implementations also became available and, by the end of 1991, the Internet has grown to include some 5,000 networks in over three dozen countries, serving over 700,000 host computers used by over 4,000,000 people. The ARPANET ceased to exist in 1990. Commercial network providers in the U.S. and Europe began to offer Internet backbone and access support on a competitive basis to interested parties. Access to Internet was first offered on commercial basis by "World" (world.std.com), thus it became the first Internet Service Provider (ISP) of Internet dial-up access. Several other countries got connected to the Internet in 1990 including Argentina, Austria, Belgium, Brazil, Chile, Greece, India, Ireland, South Korea, Spain and Switzerland.

Wide Area Information Servers (WAISs) were invented in 1991 by Brewster Kahle and released by the Thinking Machines Corporation. These servers became the basis of indices to information available on the Internet. The indexing and search

techniques implemented by these engines allow Internet users to find information using keywords across vast resources available on the net.

The most significant development in the history of Internet was the invention of World Wide Web (WWW) by Tim Berners-Lee at the CERN Laboratory in 1991. The first Web browser called "Mosaic" was released in 1993 that took the Internet by storm. Several other countries got connected to the Internet in the year 1993. The InterNIC was created in 1993 to provide specific Internet services

including i) Directory of database services; ii) Registration services; and iii) Information services.

In 1994, the Internet (ARPANET) celebrated its 25th anniversary. Internet shopping and e-commerce commenced its operation on the net. Growth on the Internet traffic became geometric, i.e. NSFNet traffic passed 10 trillion bytes/month during 1994. WWW became the second most popular service on the net (behind FTP) leaving Telnet at third place. In March 1995, the WWW surpassed FTP as the service with greatest traffic on NSFNet based on packet count.

Several traditional dial-up systems in USA including CompuServe, America Online, Prodigy began to provide Internet access for services other than e-mail, i.e. WWW, Gopher, FTP and so on.

The technologies of the decade were WWW and search engines. New technologies emerged in late 1990s, including client-based code loaded from Web servers such as Java, JavaScript and ActiveX, etc. The research and development on the Internet and related technologies continues even today.

A great deal of support for the Internet community has come from the U.S. Federal Government, since the Internet was originally part of a federally-funded research program and, subsequently, has become a major part of the U.S. research infrastructure. During the late 1980's, however, the population of Internet users and network constituents expanded internationally and began to include commercial facilities. Indeed, the bulk of the system today is made up of private networking facilities in educational and research institutions, businesses and in government organizations across the globe.

Unit-3
Components of a Digital Library

1. Introduction

Establishing digital library resources and services require a great deal of infrastructural components that are not available off-the-shelf as packaged solution. There is no turn-key, monolithic systems available for digital libraries, instead digital libraries are collection of disparatesystems and resources connected through a network andmade interoperable using open system architecture and open protocol and are integrated within one interface, currently the web interface. Use of open architecture and open standardsmake it possible that pieces of required infrastructure, be it hardware, software or accessories, are gathered from different vendors in the marketplace and integrated to construct a working digital library environment. Several components required for establishing a digital library are internal to the institutions, but several others are distributed across the Internet, owned and controlled by a large number of independent players.The task of building a digital library, therefore, requires a great deal of integration of various components (Flecker, D., 2001). Major components required for a digital library can broadly be divided into sixmajor categories mentioned above and depicted in Figure 1.
These components are described briefly in this module. However, separate modules are devoted to impart detailed information on each of the six components of digital library mentioned above.

2. Collections Infrastructure

The most important component of a digital library is the digital collection it holds or has access to. Viability and extent of the usefulness of a digital library depends upon the critical mass of digital collection it has. The collection infrastructure typically consists of two components, i.e. metadata and digital objects that a digital library holds. The metadata provides bibliographic or index information for the digital objects. Whiledigital objects are the primary documents that users are interested to access, it ismetadata that facilitates their identification, retrieval and location using variety of search techniques. Information content of a digital library, depending on the media type it contain, may include a combination of structured / unstructured text, numerical data, scanned images, graphics, audio and video recordings and other multimedia content. Different types of resources need to be handled differently in a digital library.

The libraries, irrespective of media typesthat they house, i.e. print, audio-visual or digital, are primarily responsible for identifying, selecting, organizing, preserving and providing access to diverse categories of resources to their users. The transition from traditional library to digital library cannot happen overnight in a single step, rather this transition is gradual and incremental in nature. As such, the traditional libraries are not becoming digital libraries, but are increasingly acquiring access to ever growing digital collections for their userseither by licensing of e-resources available in the market place or by its acquisition on one-time purchase and perpetual access basis. Collections in digital

libraries may also consists of datasets that are "borne digital" or existing printed documents converted into digital format through scanning. Creating virtual libraries, libraryportalsorsubjectgatewaysarealsoconsideredasanimportantdigitallibrary

collection. Collection management in a digital or hybrid library need to have pre-defined policies and practices similar to those being followed in traditional library while keeping in view the issues and complexities that are especially related to digital materials.

The current electronic publishing market consists oftraditional players such as commercialpublishers,scholarlysocieties,universitypressesofferingelectronicversions of their print journals as well as several new enterprises offering new products and services that are "borne digital". The market also has several aggregators that provide electronic resources in a given discipline sourced from different publishers. These publishers offera variety ofelectronic resourcesincluding electronic journals, electronic books, conference proceedings, online courseware, learning materials, tutorials, guides, manuals, patents, standards, electronic e-prints (preprints and postprints), technical reports, electronic theses and dissertations, online databases and databanks, dictionaries, encyclopedias, subject portals or pathfinders. Major publishers, besides offering their electronic journals are now offering electronic books either directly through their Web sites or in partnership with other publishers or through aggregators like e-brary, NetLibrary, Questia, 24x7, Knovel, etc. Moreover, more than 32,000 books are available free of cost through Project Gutenberg. These electronic resources are available on variable pricing model.
A separate module is devoted to collection development, selection, acquisition, licensing and management of digital resources.

3. Digital Knowledge Organization

Traditional library consists of physical objects such as books, journals, conference documents, standards, patents, video, microfilms and CDs that are organized into various collections such as Text Books, General Books, Reference Books, Rare Books, Audio-visuals, CD ROM Collections and Journals. Each collection is further organizedusing classification schemes such as Dewey Decimal Classification, Library of Congress Classification, Universal Decimal Classification, Colon Classification, etc. so as to bring books on same subject together and facilitate browsing documents on the shelves. Moreover, each book is catalogued and assigned subject headings using standard subject headings and thesauri like Library of Congress Subject Headings (LCSH), Medical Subject Headings(MeSH), Sear's Subject Headings, etc. so as to facilitate their retrieval using Library OPAC. While physical libraries are organized at physical level, i.e. books, journals, theses, reports,reference books, textbooks, etc., digital libraries are organized at digital objects level which may include a combination of structured / unstructured text, numeric data, scanned images, graphics, articles in a journal or chapters in a book and other multimedia objects.

A disc full of digital objects without any organization, browse, search and navigation options would be completely useless and meaningless since these digital objects need to be organized and made accessible to the user community. An effective and efficient access mechanism that allows a user to browse, search and navigate digital resources becomes necessary as electronic resources of a collection grow in number and complexion.AsdigitallibrariesarebuiltaroundtheWebandInternetTechnology,ituses object

and addressing protocols of the Internet. The process of organizing digital objects includes: i) development of metadata schema; ii) assigning different kind and levels of metadata to each digital object; iii) assigning Unique Object Identifiers to each digital object; iv) linking digital objects with associated metadata to facilitate their browsing, searchingandnavigation;andv)organizingdigitalobjectsandassociatedmetadataintoa database; and vi) building browse, search and navigation interfaces.

A separate module is devoted to digital knowledge organization that would deliberate on the issues mentioned above.

4. Access Infrastructure: Browse, Search and Navigation Interfaces of Digital Library

An effective and efficient access mechanism that allows a user to browse, search and navigate digital resources becomes necessary as electronic resources of a collection grow in number and complexion. While the access infrastructure for a traditional library is OPAC/Web OPAC (including journals holding), the access infrastructure for digital libraries consists of browse, search and navigational interfaces for individual digital libraries, specialized indices for specialized local collections, portals or subject gateways for web resources and an integrated interface for all e-resources accessible to a given library including library OPAC.

Search, Browsing and Navigational Interfaces

The users interact with the digital library using its search interface which typically supports browsing, searching and navigation. The search interface provides a visual window for users to search and browse relevant information stored in a digital resource and to display it. Most digital libraries support searching with varying degree of capabilities ranging from "simple search" to "advanced search". In the simple search mode, a user is required to enter his or her "query" in the search box. In the advanced search mode, a user can use Boolean queries, wild cards, phrase searches and field-specific searches. Many digital libraries also support relevant-ranking of search results, based on the relevance score of the retrieved documents. A typical digital library implementation may employ a variety of information retrieval techniques including metadata searching, full-text document searching and content search or combination or two or all of them. Digital libraries consisting of images also support image-based searching based on the names of objects appearing in the images. Digital libraries built around Geographical Information Systems (GIS) with Geo-spatial data, support retrieval of the relevant portion of maps which can be zoomed-in or zoomed-out (Example: National Geographic Map Machine: http://plasma.nationalgeographic.com/mapmachine). Information retrieval in a digital library is made more effective and user's-friendly by preprocessing digital documents to extract additional metadata before storing them in a database. The database is then configured to generate indices from selected fields including authors, titles, abstracts, etc. or it may also be configured to generate indices from the full-text articles with a pre-defined stop-word list. Depending upon the implementation of digital libraries, the search conducted maybe restricted to a single server or several servers geographically dispersed at distant locations. Digital libraries also support "federated search" wherein the search query is sent to search systems on

different servers and results received from different servers are merged and presented to the user. A Typical example of "federated searches" is Networked Digital Libraries of Theses and dissertations (NDLTD) project (http://www.ndltd.org/).

Besides search interface, a browsing interface is a necessity for a digital library to give a user a sense of the amount and variety of material and the attributes of these materials available in the digital library. Browsing helps a user to learn about the collection in general, topics covered and the kinds of material available in a digital collection (Marchionini, 1998). The browsing interface ofdigital library generally consists of combination of hierarchical menu and selection buttons, where the interface guides the user, starting from the top-level subject category through a series of progressively narrowing levels within the category for a user to select and retrieve associated digital objects from the digital library.Browsing interface for a full-text library, for example, may consist of research articles arranged alphabetically byi) author's name, article title and year of publications as a selectable criteria; and ii) hierarchical presentation of research articles under subject categories. Most digital library support browsing facility through the table of contents which are linked to their full-text or to the specific chapters and sections.

Digital libraries not only consist of multitude of resources, but also multitude mechanism to access these resources. A number of standards and technologies are now available that enables interoperability and cross-searching of digital libraries. Two approaches that are used for implementing cross-searching of multiple heterogeneous digital repositories includes i) metadata harvesting approach, also called discovery services; and ii) a distributed or federated searching approach that provides direct, real-time access to information sources on the web without resorting to crawling or replicating or harvesting metadata. Examples of commercially available federated search solutions include 360 Search and MetaLib (Serials Solution), Knimbus (GIST), Primo Central (Ex Libris) and EBSCOhost Discovery Service (EBSCO) and Encore Discovery(Innovative Interfaces) are examples of discovery services.

Separate modules are devoted to access infrastructure, designing browsing, search and navigational interfaces, federated and discovery search solutions, etc.

5. Network and Computing Infrastructure

A typical digital library in a distributed client-server environment consists of hardware and software components at server side as well as at the client's side Clients are

machines that are used for accessing digital library by users while the server hosts databases, digital objects, browse, search andnavigational interfaces to facilitate its access.

Computer hardware, software and network infrastructure for a digital library can broadly be divided into the following fourcategories:

i) Server-side Hardware Components including input devices, storage devices, Communication Devices, etc.;

ii) Server-side Software Components including image capturing or scanning software,

image enhancement and manipulation software, web servers, information retrieval software, Optical Character Recognition (OCR) software, Database Management System (BDMS) Software, Digital Rights Management (DRM), etc.;

iii) Client-side Hardware PCs, laptops and mobile devices and

iv) Client-side Software Components including Web browsers, Adobe's Acrobat Reader, media players, word processing software, spreadsheet software, image processing software, etc.

A separate module is devoted to deal with computer hardware, software and network infrastructure requirement of a digital library.

6. Intellectual Property Rights (IPR) and Digital Rights Management

a. Intellectual Property Rights (IPR)

Copyright has been called the "single most vexing barrier to digital library development" (Chepesuik, 1997). The current paper-based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects are less fixed, easily copied, and remotely accessible by multiple users simultaneously. The libraries, unlike private businesses or publishers that own their information, are simply caretakers of the information. Physical ownership or possession of material by a library is not necessarily an indicator of ownership of corresponding copyright. It is unlikely that libraries will ever be able to freely digitize and provide access to the copyrighted materials in their collections. Instead, the developers of digital libraries are obliged totake permission for inclusion of copyrighted material in digital form or develop mechanisms for managing copyright mechanisms that allow them to provide information without violating copyright. Copyrights and IPR issues are governed by the constitutions of various countries and through international treaties like the Berne Convention.

"Fair Use" is an exception to copyright protection that permits limited use of copyrighted material without explicit permission of the owner for non-commercial and non-profit educational purposes. Protection and ownership of intellectual property in the age of Electronic information are especially confusing in light of traditional copyright laws.

Discussions are taking place at various platforms to review the existing copyright laws in the light of electronic information. Since the images are electronically forwarded around the Internet, it becomes very difficult to control and define what can and cannot be done.

Copyright is manifested in terms of licenses and agreements in the digital world. A library is required to sign licenses to acquire access to a digital collection. The terms of licenses for digital collection varies in terms of conditions, the variety of pricing models and access limitations(see Collection Development – licensing contents). The library associations and publishers are working on model licenses that can be adopted uniformly. The libraries can negotiate with the publishers on behalf of their institutions or as a consortium of libraries.

A Conference on Fair Use (CONFU) in January1996 working party comprising both publishers and Librarians began the process of developing practicable guidelines for fair

use of electronic information. The first discussions concerned the scanning and storage, reproduction and distribution of materials in an electronic preservation system. The working party failed to agree on any guidelines but the dialogue is still alive and is expected to result in some guidance to both libraries and academics on what is permissible without prior permission (Cox,1997).

Section 108 of US Digital Millennium Copyright Act (DMCA-2000) gives libraries the right to archive upto three copies of unpublished or published materials owned by them for preservation or security purposes as long as copies are made to replace a copy the library has or used to have in its collection that has been damaged, deteriorating, lost or stolen, or the format has become obsolete. Such published works must also be out of print. The items being preserved can be in any format (text, images, sound, etc.). Furthermore, the copies can be digital, so long as they are not distributed digitally nor made available to the public in a digital format outside the premises of the library or archives.

Digital Rights Management and Access Control in Digital Library

Access management variably called, access control, terms and conditions, licensing conditions and Digital Rights Management (DRM) refers to control of access to digital libraries. Digital Rights Management (DRM) is a system of solutions created or designed as a means to prevent unauthorized access, duplication and illegal distribution of copyrighted digital media. The DRM technology was created for the publishers as a means to stop illegal reproduction and distribution of their products. In online environment,thescopeofDRMcanbeleveragedtocontrolaccesstoandusageofdigital objects and to impose restrictions on their misuse.

Four distinct aspects of access management are: i) license agreements and policies; ii) userauthenticationandauthorization);iii)accuracyandintegrityofdigitalcontent;and
iv) accessibility including permissions to operate on digital objects or its metadata. License agreements and policies are negotiated between the publishers and librariansor consortiacoordinatorsforprovidingaccesstodigitallibraries.Usersareauthenticated

and authorized to access content of a digital library as per the terms and conditions of the license agreement. While users, duly authenticated, are allowed access to information accordingtotheirnatureofclearancesandauthority,unauthorizedusersareblockedfrom accessing information. Confidentiality is of paramount importance in digital libraries containing national defence information or highly proprietary information. Accuracy or integrity means the continuing integrity of information stored in digital object servers. Digitallibrarymustnotallowaccidentalorintentionalcorruptionofinformationstoredin it by unauthorized users or programs. Accessibility means that a secure computer system must keep information available to its users. The hardware and software of a computer system should keep working efficiently and the system should be able to recover quickly in case of disaster. Moreover, users are given access to digital contents with permissions to download (in case of users) and to add, edit, delete or amend in case of editors(Russell, D and Gangemi, G.T., 1991).

It is not only essential to ensure security of data on servers and clients but also during communication between clients and servers and vice versa to ensure authenticity and integrity of data. It is possible for a hacker to eavesdrop on communication between a user's browser and a Web server and hack sensitive information, such as a credit card number, login ID and passwords or any other confidential data. A hacker could try to impersonate authorized users in order to get information which is normally not disclosed without authorization. Incidences of hackers getting access to important Web sites and defacing them are not uncommon. Techniques of data encryption are used for communicating sensitive information such as User's password and PIN codes.Encryption renders data unintelligible and unusable even if accessed by an unauthorized person. Digital certificates are deployed to establish secure communication between clients and servers.

UserAuthentication

A combination of one or more of the authentication mechanisms are deployed by the publishers for allowing access to the digital content to the authorized users hosted in digital libraries. These authentication mechanisms are: i) Log-in ID and Password-based Access; ii) IP Filtering; iii) Web Cookies; iv) Web Proxy; v)Athens; vi)Shibboleth;and vii) Referring URL. These authentication mechanisms are described in detail in the module on access management.

UserAuthorization

The processofauthenticationascertainstheidentity ofauser,whileauthorizationdefines his or her permissions in terms of access to e-resources and extent of its usage. Authorizationisgrantedtothesuccessfullyauthenticateusersaccordingtohis/herrights information available in the Access Management System (AMS). A user dulyautheticated by one of the authentication mechanism described above may actually be entitled to access only a portion of digital collection subscribed by his / her institution. Forexample,anauthenticatedusermaybeauthorisedtoaccesselectronicjournalsfroma publisher'swebsitebutnotelectronicbooks,referencesourcesorotherresources

depending on what his institution has subscribed to. Typically, all users in an institution are authorized to access all the subscribed e-resources. However, it is possible to define different levels of authorization for different categories of personnel in an institution. Besides, authorizing users of a digital collection, authorization also addresses the issue of responsibilities assigned to different personnel involved in the development of a digital library and their respective authorities in terms of addition, deletion, editing and uploading of records into a digital library. Personnel involved in the development of a digital library may be assigned different levels of authority. Authorization is more challenging than authentication, especially for widely distributed digital libraries. Access control is one method for enforcing authorization. Typically, it assumes that the user or entity has already been authenticated. Access control policies that are in vogue include i) Mandatory Access Control (MAC); ii) Discretionary Access Control (DAC);iii) Role Based Access Control (RBAC); andiv) Content Dependent Access Control (CDAC).These access control policies are described in detail in the module on access management.

Technology of Access Control and Access Tracking in Digital Library

A number of copy-protection and access control technologies have been devised that would either restrict or completely stop unauthorised use of copyrighted digital material. These technologies include i) Digital Watermarking; ii) Control on Extent of Use and Subsequent Use; iii) Fractional or Partial Access; iv) Flickering; v) and Digital Object Identifier (DOI).These access control technologies are described in detail in the module on access management.

A separate module is devoted to deal with Intellectual Property Rights (IPR) and Digital Rights Management (DRM).

7. Digital Library Services

The library research and development in digital libraries, in the beginning, was focused mainly towards providing search and browsing interface to its collection. However, providing access to its resources is only one of the several services offered by a traditional library to its users. Reference services, for example, provide personalized services to a user with human touch. The importance of reference service has increased many-fold with introduction of new information technologies in libraries. Users, who are not well versed with use of web and Internet technology, find it difficult to retrieve information from plethora of resources accessible to them from various digital repositories. Sloan (1998) emphasised that technology and information sources, on its own, cannot make up an effective digital library. Helping users in finding resources, either in physical or electronic environment, is the foremost task of a librarian.

The digital resources and associated technical infrastructure is only a means to generate services keeping its potential users in mind. Like printed resources are used in traditional libraries to generate services by the library staff, the digital resources areused to generate servicesusingsoftwaredrivenweb-basedinterfaces.Computerprogramssubstitutefor

the intellectually demanding tasks that are traditionally carried out by skilled professionals. Activities that require considerable mental activities, like reference service, cataloguing and indexing, seeking information, etc. are performed by computer programs through web-based interface with or without human interventions.

Web-based digital resources can potentially support a range of traditional and non-traditional library services. While it is recognized that librarians may not be responsible for the design and implementation of digital library infrastructure, they, as managers of digital libraries, are responsible for generating and creating awareness about digital library-based services. Most of the library services generated using digital resources resemble closely to those generated manually with improvements and modifications to suit the requirements of automated services. However, digital resources have also been used to generate innovative services that did not have a counterpart in manual parlances. While a separate module deals with digital library services in detail, these services are mentioned briefly here.

a. E-mail Alerts

The service, variably called as E-mail Alert, Table of Contents Alert, News Alert, etc., offer the ability to set up an e-mail alerts for the table of contents from a specific journal or group of journals by the end user. A user can subscribe to e-mail alerts to get periodic emails with links to new content automatically that are added to the publisher's web site. Theservice,offeredbymostofthedigitallibrariesanddatabases,canbroadlybeequated to Current Awareness Services (CAS) offered by traditional libraries.

The first time when a user requests an e-mail alert or table of contents alert, he or she is required to create a personal user profile / user login. A user is prompted to provide details such as name, email address, postal address, field of interest, user name, password, etc. Once these details are filled-in and a login ID and password is assigned to the user, he/she is required to login on to the publisher's website and the in from there he / she can start creating his / her user profile. A user may select journal titles or subject areas that he / she would like to receive regular email alerts for. All e-journal publishers that provide an email alerting service, provide some kind of on-line help and /or FAQs. Publishers offer a variety of email alerts, including ToC alerts, new issue alerts, citations alerts, publications alerts, online first alerts, search alerts, favourite journals alerts, etc.

b. **Web Feeds: RSS Feeds or Atom**

Webfeedsaredataformatsusedforprovidinguserswithfrequentlyupdatedcontent.The two main web feed formats are RSS and Atom. RSS stands for Real Simple Syndication or Rich Site Summary and Atom format was developed as an alternative to RSS. The technology, on one hand allows a web site to list the newest published updates (like table of contents of journals, new articles) through a technology called XML, on the other hand, it facilitates a web users to keep track new updates on chosen website(s). Like a personal search assistant, RSS feed readers visit pre-defined websites, look for updated

information and fetch it automatically on to the user's desk top. In order to use RSS Feed, users are required to download RSS feed reader or RSS feed aggregator, which can be web-based, desktop-based, or mobile-device-based and then "subscribe" to the RSS feeds by copying a link from the web site of a digital repository into their feed reader. The reader can then check the subscribed feeds to see if any of those feeds have new content since the last time it was checked, and if so, retrieve that content and present it to the user. Both RSS and Atom are supported by most of the feed readers.

Digital repositories of most of the publishers provide RSS Feed for delivering the contents of their journals to their users. RSS feeds on web pages are typically represented by a rectangle with the letters 

or 

. Users generally have a choice to get all the contents of issues of a journal or get contents on a given topic or subject.

c. Ask-An-Expert

"Ask-An-Expert" is a service offered by several digital libraries and databases. It is an Internet-based question and answer service that connect users with experts who possess specialized subject knowledge and skill in a given domain. Digital libraries and database provides this service as platform to connect users with experts who can answer specific question and instruct users on developing certain skills. This service is often restricted to a given user community and subscribers of a database or full-text resource.

d. **Electronic Document Delivery Services**

The term "electronic document delivery systems" implies delivery of the electronic version of a document that might involve reproduction of an electronic copy of a document if it is not already available in electronic format. However, with availability of most of the peer reviewed research journals in electronic format, most publishers and aggregators facilitate online electronic document delivery services that allow a user to download an article in full-text from their site at a pre-determined cost. Different publishers and aggregators have offers different payment options, i.e. some charge each time the journal is used, whereas others provide restriction-free access for an annual subscription.

e. Web-based User Education

Publishers of full-text electronic resources and bibliographic databases offer web-based user's guides, teaching tools and "spoken tutorials" facilitating users to make best use of resources made available by the publisher. These guides may include colour graphics, screenshots and animations. The web-based user education provides a high degree of interactivity and flexibility to the users, offering them the benefit of self-pace, graduated to teach from basic to highly advanced levels and designed in a wide range of formats that accommodate diverse learning styles. The proliferation of web-based full-text e-resources and databases has generated greater demands for such reference and instructional services. With the availability of digital resources that can be used anywhere, at any time, requirement for instructional and reference services has also grow/grown.

f. Digital Reference Service

Digital Reference Service, also called Virtual Reference Service or "Ask-A-Librarian" is a service wherein libraries or a similar voluntary organization offer reference services to theuserstypicallythroughe-mailorviaweb.Itisgenerallyconsideredasanextensionof library's existing reference service to their users who are not able to visit the library in person. In case of voluntary organizations offering digital reference services, people who serve as digital reference experts (also called volunteers or mentors) are most of the time information specialists,affiliatedtovariouslibraries.Mostdigitalreferenceserviceshave a web-based question submission form or an e-mail address or both. Users may submit questions by using either form. Once a question is read by a service, it is assigned to an individual expert for answering. An expert responds to the question with factual information and or a list of information resources. The response is either sent to the user's e-mail account or is posted on the web so that the user can access it after a certain period of time. Many services have informative web sites that include archives of questions and answers and a set of FAQs. Users are usually encouraged to browse archives and FAQs before submitting a question in case sufficient information already exists.

The Question Point service (www.QuestionPoint.org), is a subscription base service that provides libraries with access to a growing collaborative network of reference librarians in the United States and around the world. Library patrons can submit their questions at any time of the day or night through their library's Web site. The questions will be answered online by qualified library staff from the patron's own library or may be forwarded to a participating library around the world.

g. Real time Digital Reference Service: Library Chat Rooms

Many libraries are experimenting with Internet chat technology as an innovative method for offering real time digital reference service, using chat software, live interactive communication software, call counter management software, web contact software, bulletin board services, interactive customer assistance system, etc. While digital reference service is asynchronous method of information delivery, the Internet chat providing the benefit of synchronous communication between a user and a reference librarian (or mentor).Interactive reference services facilitate a user to talk to a real, live reference librarian at any time of day or night from anywhere in the world. Unlike with email reference, the librarian can perform a reference interview of a sort by seeking clarifications from the user. The librarian can conduct Internet searches and push websites onto the patron's browser, and can receive immediate feedback from the patron as to whether his or her question has been answered to his satisfaction. Several institutions in US including Cornell University, Internet Public Library, Michigan State University, North Carolina University are offering Internet chat-based service using software like Live Person, AOL Instant Messenger, Conference Room and Google Talk.

LiveRef(sm)(http://www.public.iastate.edu/~CYBERSTACKS/LiveRef.htm)maintains an online registry of real-time digital reference services.

h. My Settings, My Saved Searches and My Saved Articles

These features facilitate a user to define his or her preferences, to save and retrieve previously saved search strategies and to view articles that have been saved in previous sessions.

Discussions are taking place at various platforms to review the existing copyright laws in the light of electronic information since the images are electronically forwarded around the Internet it becomes very difficult to control and define what can and cannot be done.

Copyright is manifested in terms of licenses and agreements in the digital world. A library is required to sign licenses to acquire access to a digital collection. The terms of licenses for digital collection varies in terms of conditions, the variety of pricing models and access limitations(see Collection Development – licensing contents). The library association sand publishers are working on model licenses that can be adopted uniformly. The libraries can negotiate with the publishers on behalf of their institutions or as a consortium of libraries.

A Conference on Fair Use (CONFU) in January1996 working party comprising both publishers and Librarians began the process of developing practicable guidelines for fair use of electronic information. The first discussions concerned the scanning and storage, reproduction and distribution of materials in an electronic preservation system. The working party failed to agree on any guidelines but the dialogue is still alive and is expected to result in some guidance to both libraries and academics on what is permissible without prior permission (Cox,1997).

Section 108 of US Digital Millennium Copyright Act (DMCA-2000) gives libraries the right to archive upto three copies of unpublished or published materials owned by them for preservation or security purposes as long as copies are made to replace a copy the library has or used to have in its collection that has been damaged, deteriorating, lost or stolen, or the format has become obsolete. Such published works must also be out of print. The items being preserved can be in any format (text, images, sound, etc.). Furthermore, the copies can be digital, so long as they are not distributed digitally nor made available to the public in a digital format outside the premises of the library or archives.

1. NetworkandComputing Infrastructure

A typical digital library in a distributed client-server environment consists of hardware andsoftwarecomponentsatserversideaswellasattheclient'sside.Clientsare

machines that are used for accessing digital library by users while the server hosts databases, digital objects, browse, search andnavigational interfaces to facilitate its access.

Computer hardware, software and network infrastructure for a digital library can broadly be divided into the following fourcategories:

v) Server-side Hardware Components including input devices, storage devices, Communication Devices, etc.;

vi) Server-side Software Components including image capturing or scanning software, image enhancement and manipulation software, web servers, information retrieval software, Optical Character Recognition (OCR) software, Database Management System (BDMS) Software, Digital Rights Management (DRM), etc.;

vii) Client-sideHardwarePCs,laptopsandmobiledevices;and

viii) Client-side Software Components including Web browsers, Adobe's AcrobatReader, media players, word processing software, spreadsheet software, image processing software, etc.

A separate module is devoted to deal with computer hardware, software and network infrastructure requirement of a digital library.

Unit-4
Search and Browse Interface in Digital Library

1. Introduction

In case of traditional libraries, the role of a librarian is to enrich and organize the library collectionsso as to enable users to locate individual items easily. Likewise, the primary aim of a digital library is to guide its users to quickly identify the most reliable and suitable digital items whether they are stored in its own collection physically or to provide access to content from other digital libraries located at a remote location.In traditional library, the users search for their desired information sources by searching the physical catalogue or the online catalogue (OPAC) in case of automated libraries and locate them on shelves. But in case of digital libraries, the digital items may be available on a number of servers distributed on different physical locations.The interface of a digital library is the sole medium which directly interacts with the users for discovering the collection and content stored in a digital library and catering the information needs of the users. As Galitz (2002) defines it, "the user interface is the part of a computer and its software that people can see, hear, touch, talk to, or otherwise understand or direct."Interface designs are guided by an assessment of user needs and aimed to maximize interaction with primary resources and support both browsing and analytical search strategies.

2. Layout & Designing

Elaborating a conceptual model for interface design Arms(0)[1]emphasizes that interface design encompasses what appears on the screen and how the user manipulates it; among its aspects are fonts, colors, logos, keyboard controls, menus, and buttons. Functionally design specifies the functions that are offered to the user. Typical functions include selecting parts of a digital object, searching a list or sorting results, obtaining help, and manipulating objects that have been rendered on the screen. These functions are made possible by the data and metadata that are provided by the digital library, and by the underlying computer systems and networks. Conceptual model of interface design, therefore consists of the followingcomponents:

Conceptu almodel	Interface design
	Functional design
	Data and metadata
	Computer systems and networks

Dillon (2002)[2] lists five questions that designers of digital library interfaces should be addressing:

- How do we attract users to our resources, and make them stay?
- What will bring a user back to our resources again?
- How do I build an interface that supports a richer comprehension or appreciation of the contents?
- What makes the material more learnable by users?
- Can novices learn from viewing an expert's construction of an information space?

Since user interface for a digital library must display large volumes of data most effectively and efficiently, a user of a digital library should be presented with a:

- One or more overlapping windows that can be resized and rearranged.
- Intuitive interface to query and retrieve large amount of data spread through a number of resources
- Ability to change the user's perspective from high-level summarized information down to a specificparagraph of a document.

Tedd (2005)[3] opines that the interfaces were designed according to the principles that users should maximize their interactions with information resources and minimize their attention to the system itself, and that both browsing and search strategies should be supported for effective and efficient use of digital content available in a digital library.

Galitz (2002)[4] formulated the following sets of design principles that he argued should be applied to the design of a user's interface:

- Interface design should be aesthetically pleasing and attractive to the eye, as interactions primarily are in the visual realm.
- Visually, conceptually and linguistically clear and unambiguous.

- Compatible with the users and the tasks to be accomplished. Moreover, it should be compatible with earlier versions of the system, or any other similar kinds of systems (in theory, this would mean that all digital library interfaces would follow a standard design).
- Comprehensible, that is, easily learned and understood.
- Configurable, that is, easy to personalize configure and re-configure.
- Consistent in the sense of look, feel and execution; the same action should always give the same result.
- Controllable by the user, so that actions result from explicit user requests, are performed quickly, and are interruptible; the user should feel that he/she is in charge of the interface.
- Direct in the ways in which tasks are accomplished; the effect of actions on objects should be visible.
- Efficient, by minimizing eye and hand movements.
- Familiar, by using concepts and language that users should know, using real-world metaphors, and building upon users' existing knowledge.
- Flexible to the differing needs of users (in terms of their knowledge and skills, experience, personal preferences, and habits).
- Forgiving of common and unavoidable human errors; preventing errors whenever possible; and providing constructive messages in case of errors.
- Predictable on the part of users who should be able to anticipate the natural progression of each task.
- Recoverable by allowing reversible actions.
- Responsive to user requests, with visual, textual or auditory acknowledgement.
- Simple.
- Transparent, so that the workings inside the computer or database remain invisible to users.

Galitz emphasizes that although these principles taken together represent the design ideal, in practice trade-offs will be required between some of the individual principles. The desire to maintain compatibility with earlier versions of the interface, for example, may clash with the desire better to meet many of the other principles; efficiency may clash with flexibility, and so on. In a nutshell, Galitz (2002) emphasizes, "the best interface is one that is not noticed, one that permits the user to focus on the information and task at hand, not the mechanisms used to present the information and perform the task." The best digital library interfaces, then, are not the ones that on first encounter impress users with the most vivid colours, the most attention-grabbing icons, or the most

intricate screen layout; rather, they are those that unobtrusively allow users, no matter what are their personal characteristics or their task in hand, to find what they are seeking quickly, accurately and with the least effort.

Sastry & Reddy (2009)[5] proposed the following principles for user interface design for digital libraries foran effective user interaction and implementation:

- Simple : The digital library user interface should be simple and straightforward so that the basic functions are easily noticeable to the users.
- Support: The digital library user interface should provide users to control over the DL; it has to enable the users to accomplish tasks using any sequence of steps that they would naturally use. It should be more on event driven rather than menu driven.
- Familiar: The user interface of digital library should be familiar to its users, i.e, the users should not require special training to perform any task.
- Informative Feedback: The user interface of digital library must provide informative feedbacks to its user during various tasks performed by the user.
- Design Dialogues to Yield Closure: Informative feedback at the completion of a series of actions.
- Prevent Errors: The user interface should not allow users to make any serious errors. Alternately the system can be designed as insensitive to errors. It should detect the user errors and offer simple, constructive and specific instruction.
- Multimedia Support: User interface of digital library should support multimedia information.
- Profile Based Support: The digital libraries my support profile creation and provide customized services based on user preferences.
- Lithe and Simple: The user interface of the digital library should be lithe and simple without having heavy and unnecessary graphics which may slow the loading of content and create disinterest among users.
- Pan and Zoom Support: It should support the basic Pan and Zoom features.
- Accuracy: It should provide as much accurate information as possible as a poor display of information, spelling errors and grammatical errors may affect the credibility of the digital library
- Efficient Searching with NLP support: Digital Library should provide efficient search mechanisms with the excellent search interface. It should provide natural language analysis and processing techniques for effective and user-

friendly searching.

- Support of semantic approach and Resource Description Frame Work (RDF) Technologies
- Sharing and Reusing of Information: It should support sharing of the content by various mechanisms.
- Multilingual Support: It should provide multilingual support for searching and displaying regional content.

Platform Independent: The user interface should be platform independent and work effectively in all environments.

- Future Plug-ins Support: It should support plug-ins for future developments and interacting with othersystems.

Digital library generally enables users to provide some common actions for discovering the content and material stored in the archive. The following is a brief of the search and browse functionality available in Digital libraries.

Content Overview

It is a common practice to provide a brief overview of the content and collection with coverage of the materials available in the digital library. It can be simple text base or can be represented with visual animations and pictures.

Search Interface

The objective of any search is to encapsulate a user's information need in one or several words - the query

- and display the resulting matched items.[6] Most digital libraries provide a simple search box as available in any web-based search engine in which users can input their search query or keywords.

Simple Search/Basic Search

Simple Search, often called "Basic Search"is the most common feature available in any digital library which is the most preferred entry point for discovering its content. An example is shown in Fig 1, from the JSTOR Archive, an archive of various types of content from journal articles to books, images, etc. It gives a simple search box for making any free text search query which searches through all metadata fields including full-text content.

Some digital library provides simple search functionality along with options to restrict the search query to specific key metadata fields. An example shown in

Figure 2 for ACS publications which provides search for anywhere and in key fields, e.g. Title, Author and Abstract.

The e-thesis library of Library and Archives Canada, as shown in Figure 2, provide search for Name,Title, Keyword, Note, ISBN etc.

Most digital library allow the basic search techniques using simple search box as given below:

Boolean Searching: The Boolean search operators, AND,OR and NOT, are used to broaden or narrowthe search results.

The Wildcard (?) and Truncation (*) Symbols: The wild card (?) and truncation(*) symbols are used to create searches where there are unknown characters, multiple spellings or various endings.

Proximity Search: Proximity Search gives results that contain two or more terms that appear within a specified number of words (or fewer) apart in the database(s). The proximity operator is placed between the terms that are to be searched, e.g. "HTML" <NEAR> publishing will search for documents that contain the word "HTML" and "publishing" within close proximity of each other (either before or after),

i.e. it might fetch terms like “HTML and Electronic Publishing”, Electronic Publishing Using HTML”, or Publishing Electronic Text with HTML”.

Grouping Terms together Using Parentheses: Parentheses can be used to control a search query.Without parentheses, a search is executed from left to right. However, words enclosed in parentheses aresearched first. Parentheses allow you to define the way the search will be executed. The left phrase in parentheses is searched first; then, based upon those results, the second phrase in parentheses is searched. Detailed description of various search techniques is discussed in a separate module.

Advanced Search

Digital libraries also provide "Advanced Search" functionality for expert users that provides either:

Multiple search boxes to specify their search query/keyword in relation to specific fields like subject, creator, abstract, title, collection type, time period, geographic location, full-text etc;

i) Plain text box that allows a user to construct his/her search strategy using Boolean operators, proximityoperators, parentheses, etc.

An example for Advanced search screen of Science Direct as shown in Fig 4 shows the multiple search input box with the option to select the fields related to search string and selection option for Boolean operators between the fields. It also provides functionalities for limiting the search query to category of material(books or journals), selection of subject category and date range (publication years) of the content.

generation search features such as relevance ranking, spell checking, tagging, enhanced content, search facets. The facets can be deployed as tools to refine a large number of search results and to narrow down search strategy to specific interest of the users. Faceted search also ensures that there is no null result for the user.[7]

In a nutshell search interface for digital libraries should include the following functionalities:

- Simple search : with search option for searching on all bibliographic fields, grouping results byarchive/collections, sorting the search results on various fields
- Advanced search: with focus on searching specific fields with more complex search queries andfilter options
- Full-text Search: include indexing of full-text content of the items along with the metadata fieldsfor deeper search

Intelligent search: provide intelligent search functionality like auto suggestions, spell checkers,similar item suggestions.

Meta Search

Digital libraries are not limited to a single repository of digital objects. The Contents of a digital library comes from different digital repositories accessible through library portals and resource discovery gateways. To cater to these requirements, digital libraries should incorporate standard and popular federated search protocols for exploring the contents stored in digital libraries. Some important search and retrieval protocols for incorporating search via federated search or meta-search solutions in digital libraries are as follows:

- **Z39.50**: Z39.50 is an ANSI / NISO standard for information storage and retrieval. It is a protocol which specifies data structures and interchange rules that allow a client machine to search databases on a server machine and retrieve records

that are identified as a result of such a search. Z39.50 protocol is used for searching and retrieving bibliographic records across more than one library system. This protocol is not used by the Internet search engines. It is more complex and more comprehensive and powerful than searching through http. Z39.50 has been extended to allow system feedback and inter- system dialogue. Like most applications working under client-server environment, Z39.50 needs a Z39.50 client program on one end, and a Z39.50 server program on the other end.

- The name Z39 came from the ANSI committee on libraries, publishing and information services which was named Z39. NISO standards are numbered sequentially and Z39 is the 50th standard developed by the NISO. The current version of Z39.50 was adopted in 1995 superseding earlier versions adopted in 1992 and 1988.

- **SRU/SRW**: Search and Retrieval via URL (SRU) and Search and Retrieval Web Service (SRW) are Web Services-based protocols for querying Internet indexes or databases and returning search results. The web services are two types, i.e, REST (Representational State Transfer) and SOAP (Simple Object Access Protocol). The SRW uses SOAP protocol and the SRW uses REST protocol for information retrieval. [8]

NISO Meta search XML Gateway (MXG): MXG is proposed as an alternate to Z39.50 protocol and is based on the SRU protocol. The NISO MXG is a low-barrier-to-entry method for content providers to expose their content to meta search application.

- **Open Search**: It is a way for websites and search engines to publish search results in a standard and accessible format suitable for syndication and aggregation. The OpenSearch is built on XML and supports a mechanism for telling a deep web search engine how to query it and the search results data are retrieved in a highly structured format. As such, search results are easy to process and display by a federated search service.[10]

In many cases, centralized index are also maintained by harvesting the metadata of relevant sources of information. Some major standards for harvesting of metadata which should be provided for discovery of digital library content are as follows:

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides a mechanism for repository interoperability. Using OAI-PMH, the service providers can make request to the repositories to harvest metadata. This is mostly used by the repositories and some e-journals providers also expose their content using it.

a. **OAI-PMH**: Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides a mechanism for repository interoperability. Using OAI-PMH, the service providers can make request to the repositories to harvest metadata. This is mostly used by the repositories and some e-journals providers also expose their content using it

b. **METS**: Similar to OAI-PMH in purpose and function, METS supports XML-encoded metadata harvesting, but unlike OAI-PMH, METS can harvest both metadata and object.[12]

Details of various standards including standards related to search and discovery of digital library content are discussed in a separate module.

Browse Interface

Another important aspect of information retrieval in digital libraries is the browsing process. The browsing process is to retrieve the desired item from the digital library where the detailed information of the object is not known or there is a lack of clarity in information need of user.Browsing is a vital part of the information-seeking process, allowing information seekers to meet ill-defined information needs and find new information. A common definition of browsing is an exploratory information-seeking strategy relying heavily on serendipity and being used to meet an ill-defined information need. Browsing is easily shown to be a vital part of the information-seeking process and very effective when combined with searching.[13]

Browsing can be provided by different formats with different entry points. It can be provided throughmetadata based alphabetical browsing like author, title, keywords/ subject categorization (for example the Library of Congress classification scheme), publisher, publication year etc.

Image browsers (as available from Smithsonian Libraries collection (http://library.si.edu/digital- library/art-and-design)) are effective means for retrieving information in digital libraries containing images and video galleries where search queries may not be very effective in the retrieval of the items. The

image browsers can be created in this method where users can view thumbnails for the items in a group of 10, 20 and so on.

Page browser are also a kind of browser which have an important usability in collections where digitized pages are individual images for manuscripts or journals. A typical interface of page browser can be viewed in Digital library of Villanova University (http://digital.library.villanova.edu/) which provide a thumbnail of individual pages on the left panel, enabling the user to browse between the index page and the page images to go from one image to the next and zoom in for detailed view.

A modern semantic browser approach can also be provided using semantic browsing tools which will provide self organized maps and hierarchical structure of phrases based for browsing content based on semantic data. An elaborated account of semantic web and digital libraries is presented in another module.

Display of Retrieved information

Searching and browsing involves an iterative process for discovering content of digital libraries. It may happen that the required content can be retrieved in a single search query or the user may need to query several time and refine the search results for finding what exactly the user is seeking for. In order to refine the search query and make decisions of these kinds, the needs to quickly evaluate the search results. This is where the display of the retrieved information plays a key role. The retrieved search results must be displayed with precision and brevity so that maximum search results can be displayed in a single window enabling users to select the relevancy of the content as early as possible.

An example of the search result page of Ebrary is shown in fig -8 that shows the brief details of an e-book with metadata of title, author, publisher, publication date and keyword etc. along with a cover image of the book for easy identification. It also provides an option to view the table of content (TOC) of the book on the same screen for further enhancement of the search result.

The JSTOR Archive, also provide a very informative result set display where the basic information on search results are listed with a option to view a preview of the item with abstract level display along with a view of the full-text content highlighting the searched key words for understanding the context of the keywords available in the item

Specialized Digital Libraries (Video, Audio, Images etc)

Some specialized digital libraries contain various non-textual content (multimedia contents) like still images, videos and audio, etc. These specialized collections pose different challenges for the designers. Often in visual (image) content digital libraries browsing can be an effective mechanism where thumbnails of reduces size can be displayed on any page. Sound clips can be browsed using sampling at specific time intervals. But browsing may not be an effective way of retrieving these content where the collection is very large.

Uint-5
Search and Browse Interface in Digital Library

1. Introduction

The Internet and web technology are the principle mechanism deployed in a digital library to search, navigate and deliver electronic resources across the globe. The primary objective of the digital library is to meet the information need of its users. Digital libraries have to be more and more responsive by maximizing the innovative impact of advancement in information and communication technology. Development in information and communication technology have greatly changed the way of information handling. To establish a digital library there must be an infrastructure for managing indexing and disseminating multimedia content. A scalable technical infrastructure needs to be carefully planned to meet the functional requirement of digital libraries. This module will discuss the core infrastructure elements that can handle voluminous content and other complexities of digital libraries.

Hardware, server allocations, databases and distribution approaches, network infrastructure and bandwidth considerations, are key in establishing the digital library as a resource that teachers, students, researchers, and the general public regard as reliable

2. Networks and Computing Infrastructure

Establishing a digital library requires a great deal of computer (both software and hardware) and network infrastructural components that are not available off-the-shelf as packaged solutions. There are no turn-key, monolithic systems available for digital libraries, instead digital libraries are collection of disparate systems and resources connected through a network, and integrated within one interface, currently the web interface. Use of open architecture, standard and protocols, however, make it possible that pieces of required infrastructure, be it hardware, software or accessories, are gathered from different vendors and integrated to construct a working environment. While some of the components required for establishing a digital library would be internal to the institutions, but several others would be distributed across the Internet, owned and controlled by a large number of independent players. The task of building a digital library, therefore, requires a great deal of integration of various components (Flecker, D., 2000).

A digital library implementation requires an enterprise-level technology solution that is scalable both in size and functionalities with built-in reliability, availability and serviceability (RAS) features (Wright, 2002). The storage capacity of a digital library should be scalable to accommodate its ever growing collection without requiring redesign and reengineering of system design as requirements grow. The use of open systems architecture provides a robust platform, digital library management solutions and development tools. Current servers from multiple vendors are being used by several digital library implementations for its scalability, and RAS features. These servers also offer high-availability features such as full hardware redundancy, fault-isolated dynamic system domains, concurrent maintenance and clustering support along with offerings for modular storage capacity that can be added incrementally.

A typical digital library in a distributed client-server environment consists of hardware and software components at server side as well as at client's side. Clients are machines that are used for accessing digital library by users while the server hosts databases, digital objects, browse and search interfaces to facilitate its access.

Server-side Hardware Components

Servers are the heart of a digital library. Server for digital library implementation need to be computationally powerful, have adequate main memory (RAM) to handle the expected work, have large amount of secure disc storage for the database(s) and digital objects and have adequate network bandwidth to meet communication requirements. A digital library may need a number of specialized servers for different tasks so as to distribute the workload onto different servers. It would require one or more object server(s) to store digital objects and other multimedia objects, an index server that maintains indices and support searching of data stored in a distributed system and last but not least a rights management system to take care of unauthorized usage and intellectual property right issues. However, for a smaller library, many distinct activities can be performed on a single server. It is important that the server is scalable so that additional storage, processing power or networking capabilities can be added whenever required.

Input Devices

Image-based digital library implementation require input devices like scanners, digital cameras, video cameras, and touch screen systems. A large range of choices are available for these image capturing devices. Scanners are available in all sizes and shapes. Flatbed scanners or digital cameras mounted on book cradle are more suitable for libraries. Details on such input devices are available in modules on digitization as well as in modules under the paper ICT Applications in Libraries.

Storage Devices

Since digital libraries require large amounts of storage, particular attention need to be given to the storage solution. A digital library would require one or more servers to store raw data (images, text, video, etc.) indices of metadata so as to retrieve information from the digital libraries in desired fashion. Digital library collections that are too large to store entirely on a disk use hierarchical storage mechanisms (HSM). In an HSM, the most frequently used data is kept on fast disks while less frequently used data is kept in nearline such as an automated (robotic) tape library. An HSM can automatically migrate data from tape to disk and vice-versa as required. Intelligent storage area networks (SAN) and Network Attached Storage (NAS) are now available in which the physical storage devices are intelligently controlled and made available to a number of servers.

Redundancy is another important storage consideration. In a system that is completely dependent on the interaction of various kinds and levels of hardware and software, failure in any one of the subsystems could mean the loss or corruption of the information object. Effective storage management thus means providing for redundant copies of the archived objects to ensure availability of documents in case of loss. A number of RAID (Redundant Array of Inexpensive Disks) models are now available for greater security and performance. The RAID technology distributes the data across a number of disks in a

way that even if one or more disks fail, the system would still function while the failed component is replaced. Digital archives may also choose to make backup copies on their own or to make arrangements for other sites to serve as backup.

Although harddisc (fixed and removable) solutions are increasingly available at an affordable cost, optical storage devices, including CD ROM, DVD ROM, BlueRay or

opto-magnetic devices in standalone or networked mode, are attractive alternatives for long-term storage of digital information. Optical drives record information by writing data onto the disc with a laser beam. This media offers enormous storage capabilities.

Communication Devices

Setting-up a digital library requires a network and communication equipment like communication switches, routers, hubs, repeaters, modems and other items required in a Local Area Network or to connect Internet. These hardware and software items are required for setting-up any network and are not specific to a digital library.

Server-side Software Components

A typical digital library requires a number of software packages to handle its highly diversified resources, activities and services. Different softwares are required to handle different components and activities of a digital library. Software required for a digital library can broadly be categories into the following two categories:

Software Required for Content Creation

A document capturing software is required for scanning legacy documents that are not available in computer-processible file. Most scanners and digital camera come with a basic image capturing software. The images captured in the process may need manipulation to enhance their quality. Software like Adobe's Photoshop or open source GIMP (GNU Image Manipulation Program) provides image enhancement features like filters, tonal reproduction, colour management, touch, crop, image sharpening, contrast, transparent background, etc. Software like ABBYY Fine Reader provides multiple functionalities like image capturing, image enhancement and OCR.

Printed text, pictures and figures captured in the process of scanning are stored in a file as a bit-mapped page image, irrespective of the fact whether 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 it can be read by humans, but not by the computers. As such "text" in a page image is not searchable on a computer. The bit-mapped pages are converted into textual files using Optical Character Recognition (OCR) software. Most document imaging softwares have OCR package in-built. However, OCR packages, such as Scan soft, Omni Page Professional and ABBYY Fine Reader, are also available as separate utilities. Acrobat Capture also has an OCR built into it. Converting material already available in digital format into PDF requires Acrobat Software Suite (or other conversion software).

Software Required for Operations of Digital Library

Like any other server, a server for digital library requires an operating system. A De factooperating system for most digital library implementation is Unix and its variants such as

Linux. As digital libraries are built around the Web and Internet technology, the server for a digital library requires a web server software like Apache's httpd or Microsoft's Internet Information Server (IIS).

Organization of digital objects with associated metadata requires an RDBMS package such as Oracle, MySQL, MS SQL, PostgreSQLor NoSQL packages like Cassandra, MongoDB etc. The database management software provides structured storage and retrieval facilities to the contents of a digital library. Further, a digital library requires a search engine connected to a DBMS to support searching of digital objects stored in it. Dspace, for example uses Apache Lucene search engine. Moreover, the contents of a digital library may have to be offered to only authorized users. The right management software such as InterTrust Systems Developer's Kit,Active Directory by Microsoft facilitates control and monitor of access to contents of a digital library.

Since a single integrated software package from a single vendor is not available, a digital library software may be a system with components added onto an open architecture framework. For example, the Dspace, a popular, open source digital library software consists of as number of software like: Web server, DBMS (Postgres or Oracle), Apache Tomcat, Apache Ant, Java, Handles and Lucene Search Engine. Some of the important digital library softwares are described briefly below.

Digital Library Software

A number of digital libraries are being constructed at present utilizing a mixture of information retrieval, media management and web server packages. All these pieces of software need to be integrated so as to present a cohesive environment and to avoid problems with growth and expansion. However, there are few software packages that attempts to provide a number functions of a digital library in an integrated fashion. Some of the important software used in setting-up a digital library are:

DSpace (www.dspace.org) was developed in partnership between Hewlett-Packard (HP) and MIT (Massachusetts Institute of Technology) and being maintained by DuraSpace foundation. Dspace, as institutional repository software, is making its mark with an increasing number of institutions around the globe installing, evaluating and using the package. The latest stable version is 4.0 and is available for download at http://sourceforge.net/projects/dspace/.

DSpace captures, stores, indexes, preserves, and redistributes the intellectual output of aninstitution's research faculty in digital formats. DSpace accepts all forms of digital materials including text, images, video, and audio files. Possible content includes: articles and preprints, technical reports, working papers, conference papers, e-theses, datasets (statistical, geospatial, matlab, etc.), images (visual, scientific, etc.), audio files, video files, learning objects and reformatted digital library collections.

Greenstone Digital Library (GSDL)is a suite of software which has the ability to serve digital library collections and build new collections. It provides a new way of organizing information and publishing it on the Internet or on CD-ROM. The Greenstone Digital Library Software is produced by the New Zealand Digital Library Project at the University of Waikato, and distributed in cooperation with UNESCO and the Humanities Library Project. It is open-source software, available from http://greenstone.org under the terms of the GNU General Public License. The New Zealand Digital Library Web site (http://nzdl.org) contains numerous example collections, all created with the Greenstone software. The Greenstone runs on Windows and Linux platform. The distribution includes ready-to-use binaries for all versions of Windows and for Linux. It also includes complete source code for the system, which can be compiled using appropriate compiler. Greenstone works with associated software that is also freely available: the Apache Web server and PERL.

GNU E-Prints is an open source digital library software package designed primarily to create institutional repositories (http://www.eprints.org/). The default configuration creates a research papers archive. With its origins in the scholarly communication movement, E-prints default configuration is geared to research papers, but it can be adapted for other purposes and content. It was developed at the Electronics and Computer Science Department of the University of Southampton. GNU E-Prints is freely distributable and subject to the GNU General Public License. The latest version is 3.3.12 and is available for download at http://files.eprints.org/. Installing the E-prints software is relatively easy. Knowledge of MySql (used as backend database), apache WWW server and Perl programming language would be helpful. Mod_perl module for Apache significantly increases the performance of Perl scripts. Complete documentation for the installation of E-prints is available on the web site (http://wiki.eprints.org).

The **CONTENTdm** from OCLC is a multimedia software suite that provides easy loading, management and access to media archives in a library. The software provides tools to assist with every phase of collection development. One can start small with a few items or CONTENTdm can handle databases with millions of objects. The CONTENTdm technology is based on years of university research and testing that have resulted in a proven set of programs.

FEDORA (Flexible Extensible Digital Object Repository Architecture) repository system (http://www.fedora.info) is an open source, digital object repository system developed jointly by the University of Virginia Library and Cornell University and now being maintained by DuraSpace Foundation. The Fedora project is devoted to the goal of providing open-source repository software that can serve as the foundation for many types of information management systems. The software demonstrates how distributed digital information management can be deployed using web-based technologies, including XML and web services. Some of the important features of FEDORA include:

- **XML submission and storage:** Digital objects are stored as XML-encoded files that conform to an extension of the Metadata Encoding and Transmission Standard (METS) schema.
- **Parameterized disseminators:** Behaviors defined for an object support user-supplied options that are handled at dissemination time.

- **Access Control and Authentication:** Although Advanced Access Control and Authentication are not scheduled until Phase II of the project, a simple form of access control has been added in Phase I of the project to provide access restrictions based on IP address. IP range restriction is supported in both the Management and Access APIs. In addition, the Management API is protected by HTTP Basic Authentication.
- **Default Disseminator:** The Default Disseminator is a built-in internal disseminator on every object that provides a system-defined behavior mechanism for disseminating the basic contents of an object.
- **Searching:** Selected system metadata fields are indexed along with the primary Dublin Core record for each object. The Fedora repository system provides a search interface for both full text and field-specific queries across these metadata fields.
- **OAI Metadata Harvesting:** The OAI Protocol for Metadata Harvesting is a standard for sharing metadata across repositories. Every Fedora digital object has a primary Dublin Core record that conforms to the schema. This metadata is accessible using the OAI Protocol for Metadata Harvesting, v2.0.
- **Batch Utility:** The Fedora repository system includes a Batch Utility as part of the Management client that enables the mass creation and loading of data objects.

Client-side Hardware & Software Components

Clients are the machines that reside on the user's desks. Planners of the digital library, therefore, need to prescribe a minimum level of hardware and software that a user would require so as to achieve efficient and effective interaction with the digital library. Most of the digital libraries require an Internet-enabled multimedia PC (or Machintosh) or a tablet equipped with an Internet Browser like Internet Explorer,Mozilla FireFox or Google Chrome as their clients. The client-side PCs may also require the following software packages (plug-ins) to download format-specific deliverables from a digital library:

Application	Software	URL
Internet Browser	Google Chrome Internet Explorer Mozilla Firefox	http://www.google.com/chrome http://www.microsoft.com/ http://www.mozilla.org
Reading PDF Files	Acrobat Reader(Adobe)	http://www.adobe.com
For Playing Audio and Video Files	Real Player VLC Player	http://www.real.com http://www.videolan.org
File Transfer Client	WS_FTP	http://www.ipswitchft.com/
Display and printing of Word, Powerpoint , Access Documents	Microsoft Office Open Office	http://www.microsoft.com http://www.openoffice.org
TIFF Images	TIFF Viewer	http://www.alternatiff.com/
Image Manipulation and Editing	GIMP (The GNU Image Manipulation Program)	http://www.gimp.org
Video Editing	Adobe Premier	http://www.adobe.com

3. Digital Libraries and Cloud Computing

Cloud computing can be understood as a way to use off-site computer processing power to replace content creation and servers that were traditionally hosted onsite. In layman's terms, this means "using Web services for our computing needs" (Kroski, 2009). Cloud computing allows content creation to be made "when data and software applications reside on and are drawn from the network rather than locally on any one workstation". By utilizing online applications, users can create and save their files online, share content, work collaboratively with others or create entire services that can all be accessed online without need of having the programs on their own computer. These online services can reduce the need for expensive software, hardware, and even advanced technical knowledge from library staff since cloud computing services are often streamlined to be very user-friendly.

The cloud computing can be advantageous and it will increase the ability of a library to try out new software without having to buy the hardware as well as being able to scale the computing power to meet the demand of users. A library can increase the quantum of cloud computing they require by contacting their vendor instead of physically acquiring new hardware to meet increased demands. This approach will be quite cost effective in terms of money and manpower. Followings are the general advantages of having a digital library on cloud:

- Compliant facilities and processes
- Cost effective
- Enterprise grade services and management
- Faster provisioning of systems and applications
- Flexible and innovative
- Flexible and resilient in disaster recovery
- Highly secured infrastructure
- Reduces hardware and maintenance cost
- Round the clock access
- Simplicity of integration
- Simplified cost and consumption model