DIGITIZATION OF KNOWLEDGE: A FRAMEWORK

*Ishtiaq H. Qureshi
&
*Rafi Ahmad Khan

ABSTRACT

Library, repository of valuable data, information and knowledge, has played pivotal role in the development of mankind since he learned the art of writing. However, conventional libraries present some problems that are very laborious and time consuming to deal with. The management of time and the speed with which a work needs to be completed has become top priority in any enterprise which needs sufficient, accurate and real time information. Initiatives have been taken round the globe to digitize the information for its inherent advantages. This paper describes a frame work to build digital library resources that would facilitate the switching over from traditional library to digital library. The framework mainly focuses on three phases: Data Capture, Data Storage and Data Dissemination.

KEYWORDS

Data Capture; Data Storage; Data Dissemination; Knowledge Management.

INTRODUCTION

Unlike previous eras, this is the fast changing age of life long learning, where society expects that people will continually develop and gain new skills and knowledge. In this setup, in order to stay competitive and productive, education, a socio-economic process, is at the heart of the development of new knowledge and skills. Sociologically also, education is becoming increasingly important for the professional and personal success. Education and knowledge is now moving beyond the traditional brick and mortar to reach everyone in a

* Ishtiaq H Qureshi, EDI, Srinagar (J&K) and Rafi Ahmad Khan, Deptt of Management Studies, University of Kashmir Sgr (J&K).

customized manner as and when required. Internet and digitization have revolutionized this all important sector by imparting flexibility in terms of time, space and distance, and also by offering many formal degrees online.

The library, historically a cornerstone of scholarly endeavour, is reinventing itself in today's networked society to meet these new demands. Instead of a building that holds books, the library is evolving into an electronic portal to a growing global collection of digital content. The doors to this virtual library are now open 24 hours a day, seven days a week, and the library's holdings come to the user when needed. The emergence of the digital library mirrors the growth of e-learning or distance learning as the virtual alternative to traditional school attendance. As the student population increasingly turns to off-campus alternatives for lifelong learning, the library must evolve to fit this new educational paradigm or become obsolete as students search for other ways to conveniently locate information resources anywhere, any time. Today's library must include sophisticated tools that make it easy to find the best information resources, delivering it to one's desktop or mobile computing device at the push of a button.

While there are still many challenges to realizing the potential of digital information, digital library technologies and practices have developed to a level where they are within the reach of every type and size of educational institution. To that end, this publication is intended to provide a framework on how to approach the development of a digital library.

DIGITAL LIBRARY

Libraries store books, journals, magazines, dissertations, research reports, case studies, newspapers etc. in a very systematic and organised fashion. The storage in organized manner is must in order to make the storage, maintenance and retrieval easy and user friendly. In traditional libraries, searching and locating of the relevant information from the huge reservoir of knowledge is a major problem faced by the users. In these libraries, the organisation of the books is usually done by arranging the books primarily by subject, title, author etc and accessed by following signs to the appropriate floor, room, bookcase, shelf, and spine-labelled book through a call number. This is

comparatively a tedious task in this space age, where people have run out of time and want to accomplish their jobs effectively and efficiently.

Digital libraries have eliminated to a large extent the inconvenience caused by the traditional libraries. As a starting point though it should be assumed that digital libraries have the same purpose, functions, and goals as traditional libraries i.e., collection, development, management, subject analysis, index creation, provision of access, reference work, and preservation of the knowledge base. Digital libraries have some definite advantages over traditional libraries which motivate or forces transition to them:

- **Physical boundary**. The user of a digital library need not to go to the library physically; people from all over the world can gain access to the same information, as long as an Internet connection is available.
- **Availability**. A major advantage of digital libraries is that people can gain access to the information at any time, night or day.
- **Access**. The same resources can be used at the same time by a number of users.
- **Structured approach**. Digital libraries provide access to much richer content in a more structured manner, i.e. we can easily move from the catalogue to the particular book then to a particular chapter and so on.
- **Information retrieval**. The user is able to use any search term bellowing to the word or phrase of the entire collection. Digital libraries can provide very user-friendly interfaces, giving clickable access to its resources.
- **Preservation and conservation**. An exact copy of the original can be made any number of times without any degradation in quality.
- **Space**. Whereas traditional libraries are limited by storage space, digital libraries have the potential to store much more information, simply because digital information requires very little physical space to contain them. When a library has no space for extension digitization is the only solution.

- Networking. A particular digital library can provide a link to any other resources of other digital libraries very easily; thus a seamlessly integrated resource sharing can be achieved.
- **Cost**. The cost of maintaining a digital library is lower than that of a traditional library. A traditional library must spend large sums of money paying for staff, book maintenance, rent, and additional books. Although digital libraries do away with these fees, it has since been found that digital libraries can be no less expensive in their own way to operate. Digital libraries can and do incur large costs for the conversion of print materials into digital format, for the technical skills of staff to maintain them, and for the costs of maintaining online access (i.e servers, bandwidth costs, etc.). Also, the information in a digital library must often be "migrated" every few years to the latest digital media. This process can incur very large costs in hardware and skilled personnel.Digital library has multifaceted applications and based on these applications it has several subjective interpretations.
- *From an information retrieval point of view, it is a large database*
- *For people who work on hypertext technology, it is one particular application of hypertext methods*
- *For those working in wide-area information delivery, it is an application of the Web*
- *And for library science, it is another step in the continuing automation of libraries that began over 25 years ago* **(Nurnberg et al., 1995).**

In fact, a digital library is all of these things and these different research approaches will all add to the development of digital libraries. The digital library, a virtual world, is a venue for developing collaborative virtual situations that support shared work and shared "social play" **(Benford et al., 2001)**. A digital library is a networked collection of electronic information resources which are diverse, dynamic and made up of different media **(Adam et al., 2000; Wang, 1999)**. Digital libraries are about information access. Forms of distributed information systems constitute a global information

infrastructure shown up as digital library, digital museum, digital archives, eCommerce, eLearning, the Internet, the World Wide Web, and eGovernment (Borgman, 2000).

Librarians see a digital library much as an electronic version of a traditional library **(Xie & Wolfram, 2002)**. A librarian considers a digital library to consist of digital collections across subject matters. A computer scientist sees a digital library as a distributed and networked information system with attending databases and information services **(Xie & Wolfram, 2002; Sharma & Vishwanathan, 2001)**. A computer scientist thinks of a digital library as a repository and interactively manipulative database system that provides access to scientifically useful data **(Lynch, 1999)**. A user sees a digital library, collection of digital services and resources, as having the same accessibility as the World-Wide Web where an information source is only a click away **(Xie & Wolfram, 2002)**.

Further along in the continuum, **(Wellman., et al, 1996)** see a digital library of the future in which software agents use principles of artificial intelligence (AI) to perform "monitoring, management, and allocation of services and resources." They define a digital library as a "community of information agents" that would retain most of the properties of the traditional library, but would perform them using intelligent software rather than human beings.

The goal of the digital library is to assist users by satisfying their needs and requirements for management, access, storage, and manipulation of the variety of information stored in the collection of material that represents the "holdings" of the library. Users may be humans or they may be automated processes acting on behalf of or in support of human needs. Users also vary and include those who are "end" users (those not involved in the management and operation of the library but rather are the customers), library operators, and information "producers" who want their material available through the library.

DIGITAL LIBRARY SYSTEMS- FRAMEWORK

A *Digital Library Systems* (DLS) facilitates digital content management by ensuring content flow from the digital library to the persons who need to know. The first fundamental step in establishing a comprehensive and effective digital library system is to decide which

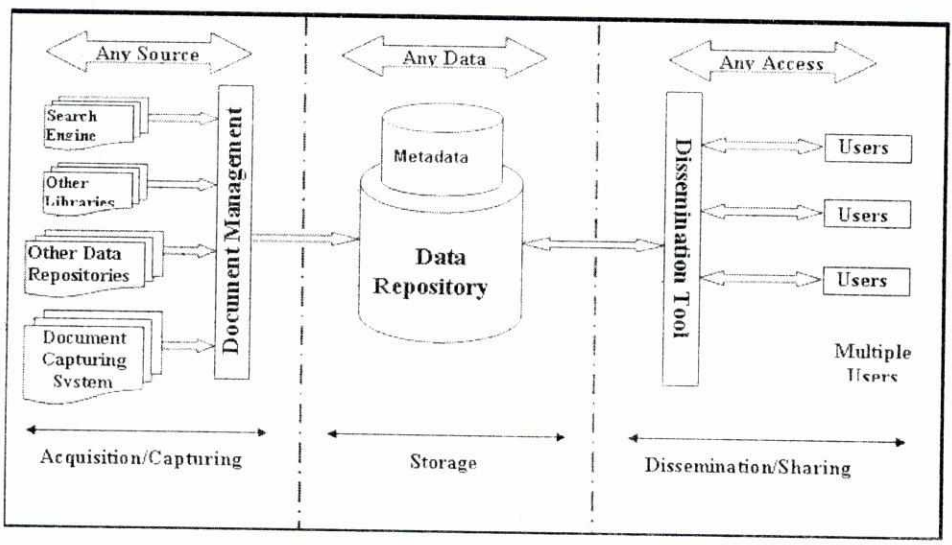
methods, tools, and technologies should be employed in implementing a digital library strategy.

Standard digital library initiatives involve the creation of digital content databases, active process management, and sharing technologies.

An attempt is made to give a clear picture depicting different components of a *Digital Library Systems*. Fig (1). These components are:

(a) Capturing/Acquisition of Digital Content
(b) Digital Content Storage
(c) Disseminating/Sharing of Digital Content

Fig 1 Digital Capture, Storage and Dissemination



Various technologies that can help in capturing of information are:

(i) Document Management System (DMS): Document Management (DM) is a critical issue for digital library, where a lot of effort is spent in properly creating, distributing and managing documents. While just some information is stored in relational databases, a relevant percentage is available in unstructured digital

formats **(the451, 2002).** Documents available in unstructured formats are commonly used text and multimedia documents. In a typical library, books, theses, dissertations, journals, reports, are available as word-processor documents, presentations as slideshows, technical seminars as a/v files and streaming media. The characteristics of unstructured documents pose several challenges for their effective management.

Document management is the art of organizing both paper and electronic flies for maximum accessibility and usefulness. Document management systems are software packages designed to help individuals, workgroups and large enterprises manage their growing number of documents stored in electronic form **(Rooney, 1992).**

Document-imaging systems convert documents and images into digital form (by scanning them) so they can be stored as image or ASCII text and accessed by a computer. Such systems store, retrieve, and manipulate a digitized document. The document is stored on a secondary storage device like hard disk or an optical disk system. Document management is used to manage the entire life cycle of a document, from creation through multiple revisions and finally into long-term storage and records management. It keeps track of masses of data and information, which is stored in a secure file vault where its integrity is guaranteed and all changes to it, are monitored, controlled, and recorded. A document management system provides far easy and faster access to all the documents. It takes care of creating, storing, editing, and distributing documents. A document management system facilitates authorization of users to specific document, along with control of the documents **(Rafi, 2003).**

Traditional document management systems can be expensive, requiring proprietary client/server networks, special client software, and storage capabilities. Intranets provide a low-cost and universally available platform for basic document publishing. Content stored on a database can be published using Web-page authoring tools and post it to an intranet Web server where it can be shared and accessed with standard Web browsers. These Weblike "documents" can be multimedia objects combining text, graphics, audio, and video along with hyperlinks. After a document has been posted to the server, it can be indexed for quicker access and linked to other documents.

(ii) Data Repositories Digital content can be acquired from data repositories of external sources like universities, government departments and organization or internal data repositories.

(iii) Search Engines These are huge databases of web page files assembled automatically by machine. There are two types of search engines:

- *Individual.*Individual search engines compile their own searchable databases on the web.
- *Meta.* Metasearchers do not compile databases. Instead, they search the databases of multiple sets of individual engines simultaneously

Search engines provide two possible services to the digital library. They allow searching the content of a site to find material, or they can allow to search on more external sites to add value to objects found on a site.

(b)Data Storage A database could be as simple as a phonebook or stock tables, or text or audio or video or as sophisticated a biological repository with terabytes of data. *Database Management System* (DBMS) is simply the software that permits to centralize data, manage them efficiently, and provide access to the stored data by applications programs. All library and information retrieval system have a DBMS underneath them. There are standard Relational DBMS like Oracle, Informix, Sybase which can all be accessed by using SQL (Standard / Structured Query Language) so the data is accessible to programs written by the library's own staff or third parties (such as Decision Support System suppliers).

There are a terabytes of digital documents and multimedia content stored all over the world in data repositories, which can be acquired by the digital library.

METADATA

Metadata by definition is simply "data about data", information about the objects stored within our collections, whether these are in traditional or electronic formats. In the standard library world, catalogue records are metadata, as they contain information about the library's collection of "data", i.e. the books and journals that make up its collections. Metadata records in the traditional library fulfill several

functions, including allowing users to find items, allowing them to assess their usefulness, and to allow librarians to administer them correctly. The same principles apply to objects within the digital library.

TYPES OF METADATA

Metadata can take several forms, some of which will be visible to the user of a digital library system, while others operate behind the scenes. The Digital Library Foundation (DLF), a coalition of 15 major research libraries in the USA, defines three different types of metadata, all essential to ensure the usability and preservation of the collection over time.

DESCRIPTIVE METADATA

This metadata provides information that allows discovery of collections or objects through the use of search tools and provides sufficient context for understanding what has been found. When collections become large or when searching multiple collections (such as over the Internet) the discovery of objects of interest becomes a "needle in a haystack" exercise. Without agreed-upon metadata standards and the discipline of capturing and storing appropriate descriptive metadata, all but the smallest digital collections would be useless.

Metadata for individual objects varies by the type of object, but would include such things as its title, what it is, who created it, contributors, language, when it was created, where it is located, the subject, etc. At the collection level, users should be able to determine the scope, ownership, any access restrictions, and other important characteristics that would assist in understanding the collection.

ADMINISTRATIVE METADATA

Administrative metadata facilitates access, management, and preservation of the digital resource. This may include information on how it was scanned, its storage format etc (often called *technical metadata*). It can describe the viewer or player necessary to access the object, automatically opening that viewer or player when a user selects that resource. It may enumerate attributes such as image resolution, file

size, or audio sampling rate. It provides copyright and licensing information, and information necessary for the long-term preservation of the digital objects (*preservation metadata*)

STRUCTURAL METADATA

This focuses on the associations within or among related individual information objects. A book, which consists of pages and chapters, is one of the most straight forward examples of structural metadata. The structural metadata would explain how individual page images make up individual chapters, and how chapters make up the book. There could also be individually imaged figures, and structural metadata could also relate these to chapters or to a list of all figures in the book. Structural metadata aids the user in navigating among individual objects that comprise a compound object.

In general, only descriptive metadata is visible to the users of a system, who search and browse it to find and assess the value of items in the collection. Administrative metadata is usually only used by those who maintain the collection, and structural metadata is generally used by the interface which compiles individual digital objects into more meaningful units (such a journal volumes) for the user.

For every library the metadata should have the following characteristics **(Smith, F., 2002)**:

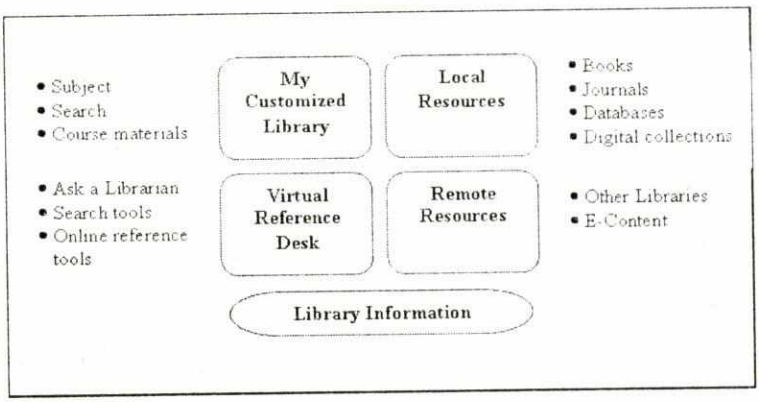
- The metadata should be kept in state-of-the-art (that is relational) databases
- The metadata should be up-to-date
- The metadata should be structured according to internationally accepted standards
- The metadata should always be connected to each other and to identifications of the material in the Data Repository
- The metadata should be correct
- The metadata should be secure and authorized
- The metadata should be relevant to all layers in the organisation
- The metadata must be created and maintained in one work process and must be available for the benefit of all work processes
- There should never be redundant metadata

(c) Dissemination/Sharing The final phase is communicating/sharing the captured content effectively in an efficient manner.

PORTALS

Deployment of portal technology has occurred side-by-side with digital collection development and access. As shown in Figure 2, portals are unified views into a set of disparate information sources and research tools, with the goal of providing users a simple way to locate and access all the information content they need and have authority to access. As libraries create, license, or negotiate access to more and more digital content, the need for an easy-to-use interface becomes increasingly important.

Fig. 2 Library Portal interfaces: An overview



Library portals typically include an online catalog of materials as well as gateways to collections of digital resources accessible to the user. Broadcast search tools allow library users to search all of these sources simultaneously with a single query. Portals may include electronic reference services ("ask a librarian"), personalization features ("my bookshelf," custom intelligent searches), and other research tools. Enriched content, such as author biographies and book reviews, tables of contents, and jacket images can be provided to supplement the online catalog. Some libraries have built interactive

features into their portals, allowing development of virtual communities.

CONCLUSION

With the advent of the Information Technology, access to the information has increased dramatically. It is no longer considered practical or acceptable to travel to a specific location during certain hours to locate needed information. The users increasingly expect instant access to all the information resources they require, from any location, at any time, and from any device. This is the objective that the digital library is fulfilling. This paper has discussed the frame work to establish digital library resources. The digitized resources would be made readily available from the desktop or other Web-enabled devices. The methods discussed are based on the technology environment available and compatible with the software tools by which these methods can be efficiently implemented. Keeping in view the scale and functionality of the digital libraries an attempt is made in this paper to emphasize on the suitability of the methods to deal with the varied requirements of the digital library initiatives.

REFERENCES

Adam, N. R., Atluri, V., & Adiwyaya, I. (2000, June). "SI in digital libraries." Communications of the ACM, 43 (6), 64-72

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001, July). "Collaborative virtual environments." Communications of the ACM, 44 (7), 79-85.

Borgman, C. L. (2001, May). "Where is the librarian in the digital library?" Communications of the ACM, 44 (5), 66-67.

Laudon K.C. & Laudon J. P. (2002). Management Information System Managing a Digital Firm, 8th ed., Pearson Education Inc (India).

Lynch, C. (1999). "Today and tomorrow: What the digital library really means for collections and services." In P. Johnson & B. MacEwan, (Eds.), Virtually Yours, (pp. 85- 97). Chicago: American Library Association.

Nurnberg, P.J., Furuta, R., Leggett, J.J., Marshall, C., and Shipman III, F.M. (1995). "Digital libraries: issues and architectures." In Proceedings of the Second Annual Conference on the Theory and Practice of Digital Libraries. Austin, Texas, June 11-13, 1995, pp. 147-153.

Rafi, Khan.(2003, Sept) "Knowledge Management: A Business Perspective." The Business Review, Vol.10, No.1, P-103

Rooney, Paula.(1992, May)."PC document management catches eye of big business", PC Week, v9 p45.

Sharma, R. K., Vishwanathan, K, R. (2001). "Digital libraries: development and challenges." Library Review, 50 (1), 10-15.

Smit, F.(2002 Feb) "The Historical Data Warehouse." Cultivate Interactive, issue 6, 11

The 451(2002). "Unstructured Data Management: the elephant in the corner". Retrieved September 22, 2006 from

http://www.the451.com/

Wang, W. (1999, June). "Information in the information age." Communications of the ACM 42 (6), 23-24.

Waters, D.J. (1998). "What are digital libraries?" CLIR Issues, July/August.
URL:http://www.clir.org/pubs/issues/issues04.html

Wellman, Michael P., Edmund H. Durfee, and William P. Birmingham (1996, June). "The digital library as community of information agents. A position statement", IEEE Expert, URL:http://ai.eecs.umich.edu/people/wellman/pubs/expert96.html

Xie, H. & Wolfram, D. (2002, November). "State digital library usability: Contributing organizational factors". Journal of the American Society for Information Science and Technology, 53 (13), 1085-1097.