

Digital Libraries: What and How?

Dr. Anil Kumar Jain

Associate Professor

S. S. In Library and Information Science

Vikram University, Ujjain



What is a Digital Library:



Digital Library - Few Definitions:

- **Borgman, CL** (1999)[\[2\]](#) point out the term digital library is used in at least in two senses:
 - i. In the computer science research community digital libraries are viewed as *content collected on behalf of users*.
 - ii. In the library practitioner community digital libraries are seen as *institutions providing a range of services in a digital environment*.

- According to **Gladney H.M, et. al.** (1994)[\[4\]](#)
“A digital library service is an assemblage of digital computing, storage, and communications machinery together with the software needed to reproduce, emulate, and extend the services provided by conventional libraries based on paper and other material means of collecting, storing, cataloguing, finding, and disseminating information.”

- **Association of Research Libraries** (1995)[\[5\]](#) has identified the following five elements in various definitions of the digital libraries:
 1. *The digital library is not a single entity;*
 2. *The digital library requires technology to link the resources of many;*
 3. *The linkages between the many digital libraries and information services are transparent to the end users;*
 4. *Universal access to digital libraries and information services is a goal;*
 5. *Digital library collections are not limited to document surrogates: they extend to digital artefacts that cannot be represented or distributed in printed formats.*

- Paul Duguid (1997)[\[6\]](#) has defined the Digital Library as an environment to bring together in support of life cycle of information in addition to digital collection and information management tools.

The concept of a "digital library" is not merely equivalent to a digitized collection with information management tools. It is rather an environment to bring together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge. (Duguid, Paul, 1997).

Digital Library Myths:

- Terry Kuny and Gary Cleveland (1998)[\[10\]](#) have tried to explode some common myths associated with Digital Libraries.

- **Myth 1: The Internet is the digital library.**

- Many different groups to signify simply a collection of digital objects that people can access from their desktops have appropriated the word “library”. A global information network, of which the Internet is the seed, has the illusion of promising fingertip access to the world’s information. A fairly spectacular example of what many people consider to be a digital library today is the World Wide Web. But is this a "digital library"? For many common library requests, locating information on the Internet remains highly inefficient compared to traditional library sources, especially for unfamiliar users. Finding information is difficult, the quality of the information is quite variable, and reliable, professional assistance for the confused and lost is lacking.

- **Myth 2: The myth of a single digital library or one-window view of digital library collections.**

- Even modest moves towards increasing digital collections and services will be strongly affected by future copyright and licensing regimes, as well as prohibitive costs for digitization and support of technical infrastructure. But more importantly, the digital future will be an unruly one composed of multiplicity of competing information providers. Libraries will be only one source of information. "Prime" information resources will probably be locked into proprietary Collections essentially "private digital libraries" which are accessible on a subscription or pay-per use basis. Developing interoperability standards for locating and retrieving information in this highly distributed and heterogeneous environment will be a considerable challenge in their own right.

- **Myth 3: Digital libraries will provide more equitable access, anywhere, any time.**
 - A great deal of work must be done to turn this myth into reality. We can assume that a global computer network or the Internet or some descendant will be the primary delivery mechanism for digital information. Equitable access is currently compromised by the fact that the Internet is not available to every one equally. Furthermore, the connections that do exist for most people are slow. For a digital library to provide equitable access to information, it is imperative that the same universal availability that is a characteristic of the telephone system is also a characteristic of the network. In the future, complex multimedia resources and services may have specialized hardware and software requirements such that only a limited number of workstations can actually access the information.

- **Myth 4: Digital libraries will be cheaper than print libraries.**

- A common assumption among technology reporters about the costs of "digital libraries" is that digital is cheaper than paper. This contention is far from established in fact or in practice. Although many libraries project savings, especially when substitution strategies are used which replace selected serials titles with document delivery services, the cost/benefit analysis of making this switch remains unclear.
- It is no surprise Digital Content providers are resorting to Contract Agreements and Licensing Mechanisms instead of normal copyright provisions.

Some Important Issues:

- **Digital Library Standards**
- **Information Resources Organisation**
- **Metadata**
- **Digital Archiving and Preservation**
- **Digital Library Services**

Digital Library Standards:

- There is a great need for adopting various standards and best practices to build interoperable digital libraries.

Dempsey *et al* (1998)[\[12\]](#) propose the following standards for use in eLib Projects:

- **User Interface:**
Common Web Browser
- **Data Handling and Interchange:**
Graphic Formats – JPEG, TIFF, GIF, PNG, Group 4 Fax, CGM
Structured Documents – SGML, HTML, XML
Moving Pictures/3-D – MPEG, AVI, GIF89A, QuickTime, Real Video, ViviActive, VRML
- **Metadata:**
Resource Description – Dublin Core, WHOIS++ Templates, US-MARC, TEI Headers, Other Open Source and Domain Specific Standards.
Resource Identification – URN, PURL, DOI, SICI
- **Security, Authentication and payment services:**
Emerging e-Commerce Standards.



Information Resources Organisation:

- **Classical Wisdom: Classification and Cataloguing**

- Classification is the process of bringing related items together. Conventional libraries to stack books on related subjects together have used modern library classification. This facilitated the browsing approaching of the information seeks and the learners. Cataloguing created document surrogates. Which provided searching facility by Authors, Titles, Series and other elements. Since classification used notation symbols, Subject cataloguing / indexing was done to provide both browsing and searching by subjects.

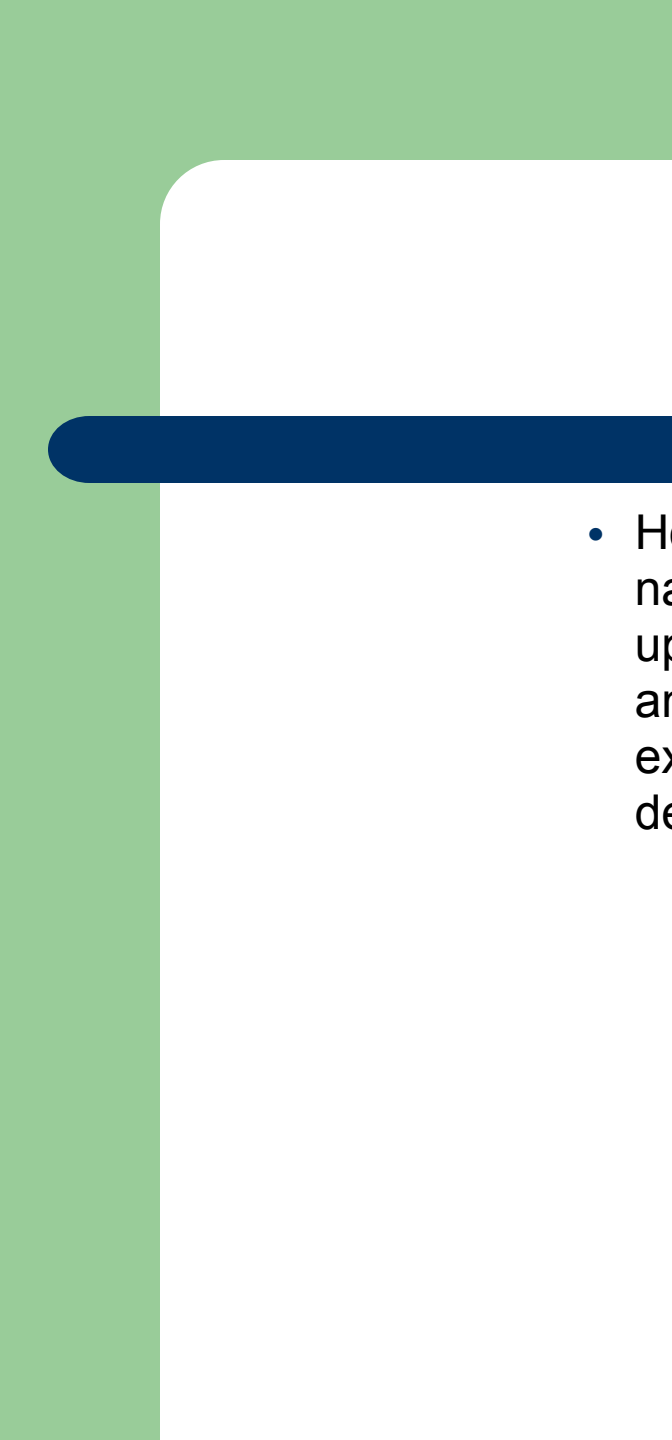
- **Problems of Information Organisation in Digital Libraries**

- Should the same classical wisdom be applied to Digital Libraries? People are debating it. Digital libraries provide access to different types of information resources, which may be organised in varying ways by their producers or access providers. It is a challenge for a given digital library to build and / or adopt a simple and yet effective method for organising information. Some researchers argue that the traditional classification schemes may not be required in digital libraries (Jones, 2002[13]), though others don't agree (E. Hunter, 2002[14]; Pollitt, 2002[15]).

- 
- 
- Traditional classification schemes are being used for organising web information resources according to the disciplines and specific subjects and topics within a discipline. Users find it comparatively easy to get information from such organised resources structures.

For example:

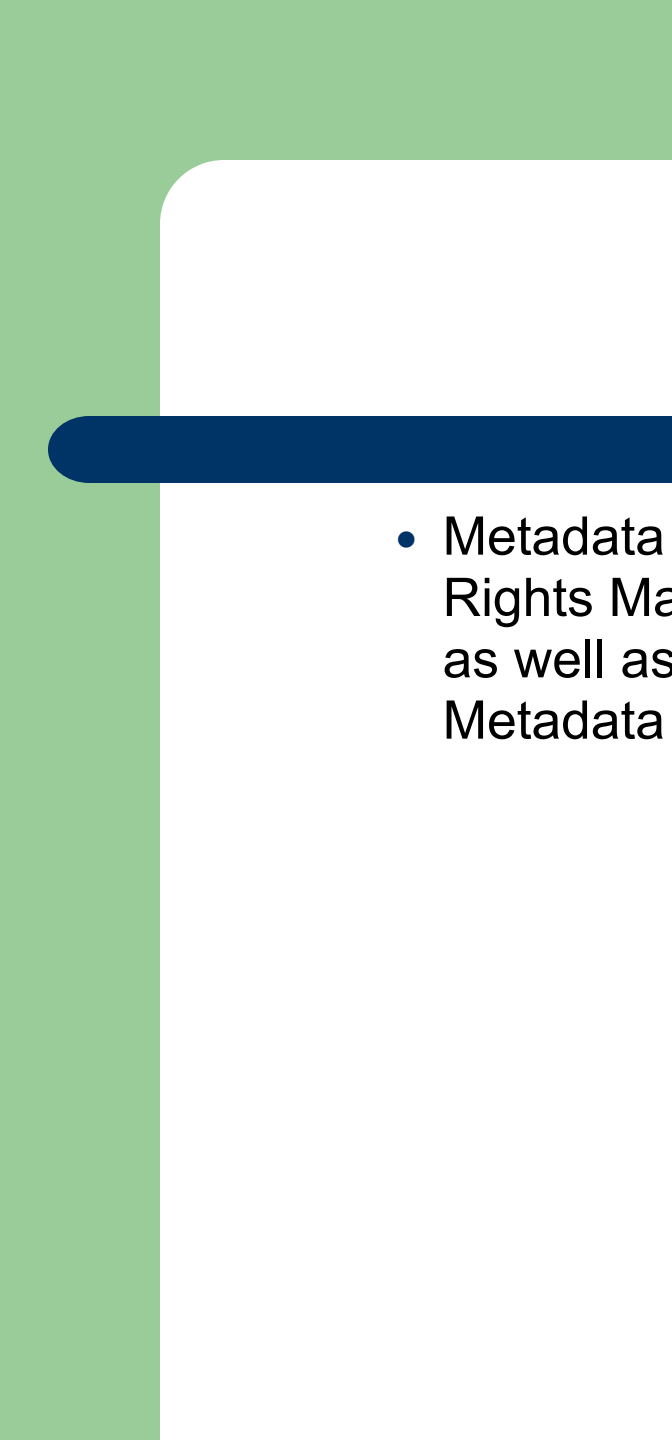
- BUBL LINK. <http://www.bubl.ac.uk/link>. DDC & LCSH
- CyberDewey. <http://www.anthus.com/cyberdewey/cyberdewey.html>.
- SCORPION. <http://orc.rsch.oclc.org:6109/>.
- CyberStackes. <http://www.public.iastate.edu/~cyberstack/>.(LC).
- INFOMINE. <http://infomine.ucr.edu>. (LCSH).
- BIOME. <http://www.biome.ac.uk>.

- 
- However, digital libraries deal with many new and nascent subjects, Classification schemes are unable to update to that pace. Moreover it is expensive to classify and catalogue according to traditional schemes by human experts. Various new metadata schemes are being developed to organise digital resources.

Metadata: What is it?

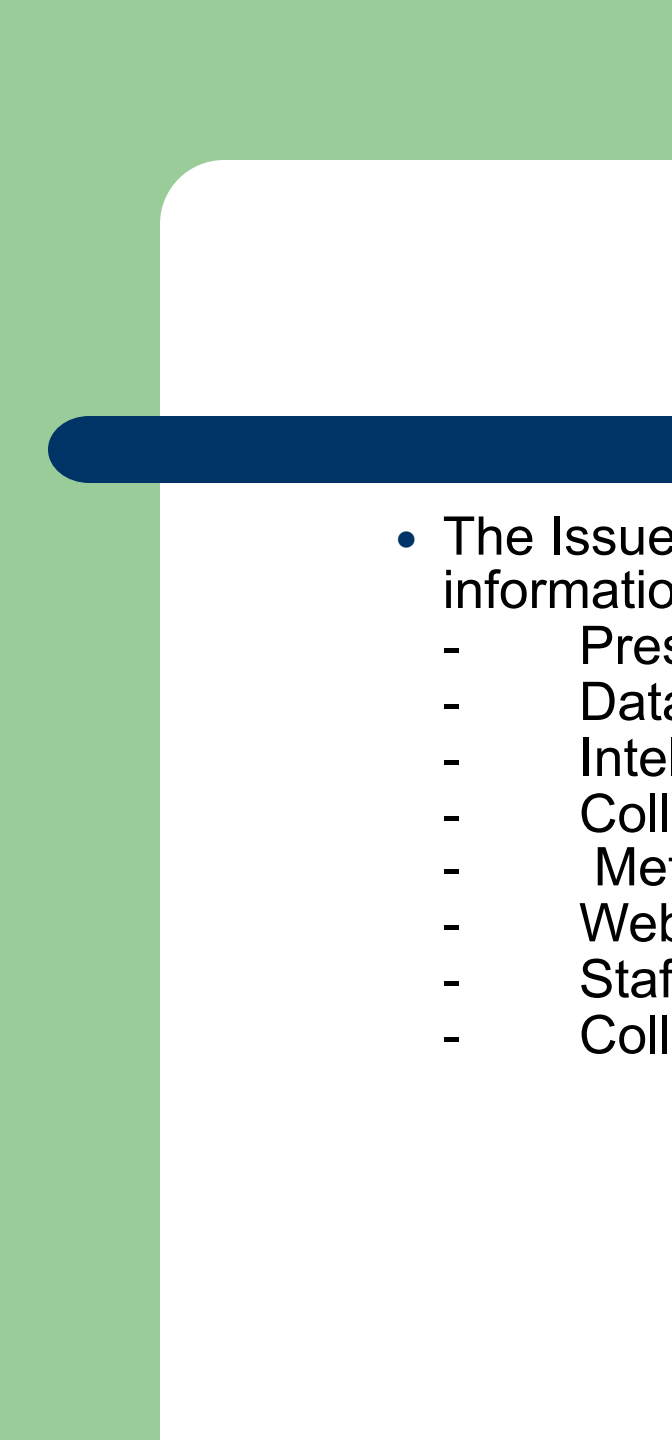
- What is metadata after all? The term entered our vocabulary with rapid growth from around 1994.
- It is nothing else but DATA ABOUT DATA.
- Vellucci (1998)[\[17\]](#) maintains that both Computer Science and Library and Information Community have adopted the term “metadata” for describing electronic data.

- Michael Gorman (1999)[\[18\]](#) argues that the non-cataloguing world perceives metadata as being different from the traditional cataloguing, which according to him, have complex formats and expansive and stringent quality requirements. He believes that four approaches are available for the bibliographic control of electronic resources of varying control. a. Full MARC cataloguing for high quality resources that are likely to have continuing value. b. Enriched Dublin Core for next level. c. Minimal Dublin Core for Next level. d. Unstructured full-text keyword searching for the reminder.

- 
- Metadata is being used for Resource Identification and Rights Management. Resource Description and Discovery as well as Preservation of Digital Resources. The use of Metadata Standards facilitates interoperability.

Digital Archiving and Preservation:

- Digital information needs to be achieved and preserved for future use. Digital information may be lost for many reasons, for example:
 - Changes in an organisation
 - Content reorganisation
 - Cession of sponsorship
 - Technology obsolescence
 - Content format obsolescence
 - Hacking and sabotage
 - Disaster, whether nature or man-made

- 
- A decorative graphic on the left side of the slide, consisting of a light green vertical bar and a dark blue horizontal bar with rounded ends.
- The Issues involved in archiving and preservation of digital information are:
 - Preservation Strategies
 - Data creators and publishers
 - Intellectual property rights
 - Collection management
 - Metadata
 - Web archiving
 - Staff expertise
 - Collaboration

Digital Library Services:

- Reference service and information services are the integral to traditional services. Although digital library research and development so far has not concentrated on them, reference services should form an important part of digital libraries.

- A number of free and fee based reference and information services are now available through web, and many of them are offered by non-library organisations.
 - Ask Jeeves. <http://www.AskJeeves.co.uk>.
 - Ask A Librarian. <http://www.earl.org.uk/ask>
 - CDRS. <http://www.loc.gov/rr/digiref>.
 - Automatic Reference Librarians for the World Wide Web.
<http://www.fastlane.nsf.gov/servlet/showaward?award=9874759>.
 - SIFTER. <http://sifter.indiana.edu>.
 - The Virtual Reference Desk. <http://www.vrd.org>

Building a Digital Library:

- **Challenges**
- **General Principles**
- **Design Principles**
- **Digital Collection Development**

Challenges:

- The staff of the National Digital Library Program at the Library of Congress (2003)[\[21\]](#) have identified ten challenges that must be met if large and effective digital libraries are to be created during the 21st century.

Building the Resource

- **Challenge One:** Develop improved technology for digitizing analog materials.
- **Challenge Two:** Design search and retrieval tools that compensate for abbreviated or incomplete cataloging or descriptive information.
- **Challenge Three:** Design tools that facilitate the enhancement of cataloging or descriptive information by incorporating the contributions of users.

Interoperability

- ***Challenge Four:*** Establish protocols and standards to facilitate the assembly of distributed digital libraries.

Intellectual Property

- ***Challenge Five:*** Address legal concerns associated with access, copying, and dissemination of physical and digital materials.

Effective Access

- **Challenge Six:** Integrate access to both digital and physical materials.
- **Challenge Seven:** Develop approaches that can present heterogeneous resources in a coherent way.
- **Challenge Eight:** Make the National Digital Library useful to different communities of users and for different purposes.
- **Challenge Nine:** Provide more efficient and more flexible tools for transforming digital content to suit the needs of end-users.

Sustaining the Resource

- ***Challenge Ten:*** Develop economic models for the support of the National Digital Library.

Principles for Digital Library Development^[22]:



Expect Change:

- Change is the only constant in the area of Digital Library. It is a fast changing world. The Technology, methods, materials and tools are changing and becoming obsolete very fast. Therefore unless you plan and take care of change, your efforts are likely to become a wasteful exercise.

Know Your Content:

- Content is the most important component of a Digital Library. Make decision to select the right content for the Digital Library. Decide what items to digitise if they exist in analog form, metadata describing content and other attributes of each item. Decide what would be the conceptual unit.

Involve the Right People:

- Involve people with variety of backgrounds that can offer expertise in building Digital Libraries. Computer and Library Professionals are mostly directly involved in the field. Within an organisation, the commitment of Senior Management is must. Lacking organisational commitment, it may not make sense to even begin a Digital Library project.

Design Usable Systems:

- Digital Libraries are mostly made available on Internet through Web technology due to the great advantage it provides. Prefer to create a simple interface avoiding server-side control of the exact display. Provide multiple access points. Consider accessibility for users with range of physical disabilities.

Ensure Open Access:

- Ensure Connectivity to the computer network and available content. Ensure that the technology is usable and the user has the requisite skills and the knowledge. Avoid proprietary hardware and software solutions whenever possible.

Be(a)ware of Data Rights:

- In conversion projects, every attempt should be made to seek permission from the copyright owner for the materials that are to be digitized. Careful tracking of permissions and privacy information ensures the collection will not be at risk at some future time.

Automate Whenever Possible:

- Building a Digital Library requires significant intellectual effort on the part of the systems creators, the more automated tools that can be built and used, the better will be the use of precious human resources.

Adopt and Adhere to Standards:

- The use of standards in system building has many advantages. It ensures more scalability, interoperability and portability.

Ensure Quality:

- Digital Library projects should define and then carry out quality control methods as part of their normal procedures.


Be Concerned About Persistence:

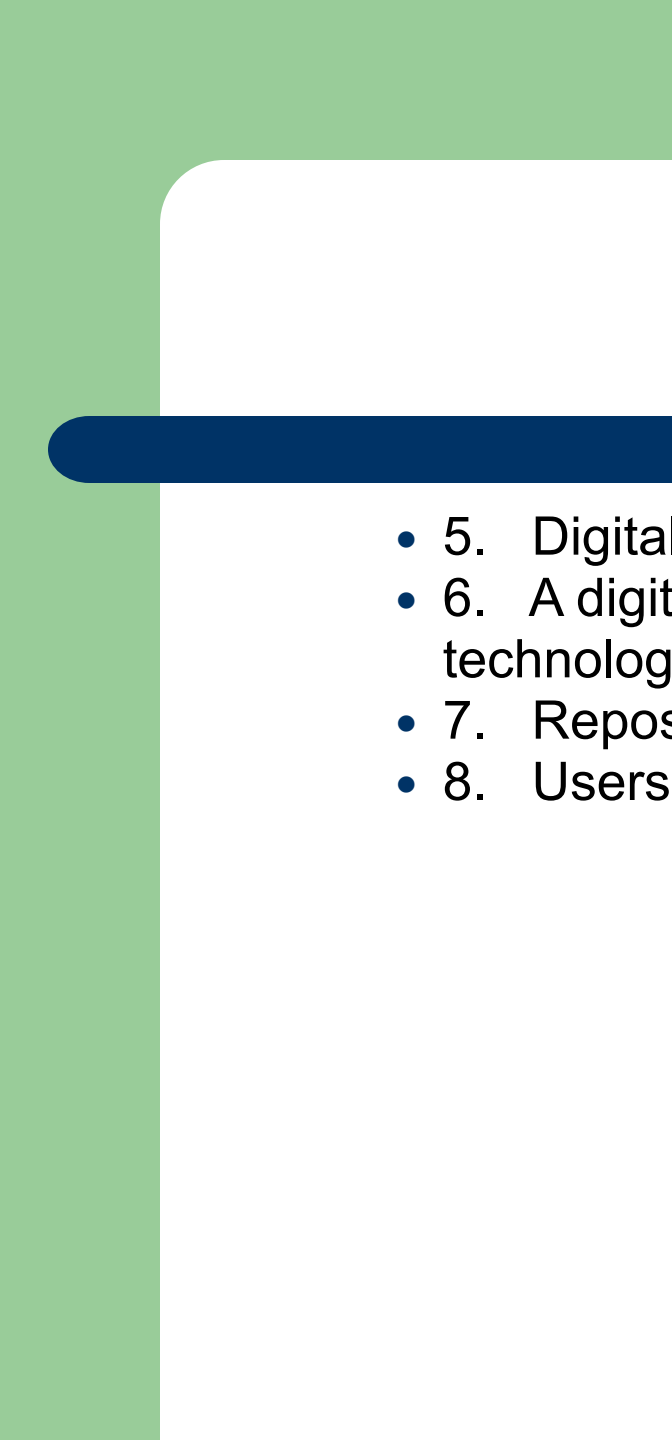
- Technological obsolescence is the greatest threat to Digital Collections. “Migration” strategy is most discussed solution. Persistence is a broader notion, encompassing preservation, but also referring to whether the object would still exist in any form at all in the future. Persistence implies a commitment to both maintaining the object and keeping it accessible.

Design Principles:

The slide features a light green background on the left side. A white rounded rectangle is positioned in the upper left, containing the text 'Design Principles:'. A thick, dark blue horizontal bar spans across the middle of the slide, starting from the left edge and extending to the right.

- Kahn and Wilensky (1995)[\[23\]](#) describe the basic infrastructure of a digital library which is “ open in its architecture and which supports a large and extensive class of distributed digital information services”. This architecture has been the subject of a series of useful discussion from which the following eight general principles have emerged (Arms, 1995)[\[24\]](#). These may be considered as the guiding principles for the design of digital libraries.

- 
- 1. The technical framework exists within a legal and social framework.
 - 2. Understanding of digital library concepts needs standard terminology.
 - 3. The underlying architecture should be separate from the content stored in the library.
 - 4. Names and identifiers are the basic building blocks for the digital library.

- 
- 5. Digital library are more than collections of bits.
 - 6. A digital library object should not be tied to a particular technology.
 - 7. Repositories must look after the information they hold.
 - 8. Users want intellectual works, not digital objects.

Digital Collection Development:



Selection of Materials:

- In conventional libraries the budget could be a limitation but here the task is much more complicated due to problem of plenty. Moreover, for identification of printed materials there are well-established book trade and selection tools like Books-in-Print, National Bibliographies, Union Catalogues etc along with the users suggestions and approvals. In case of digital content such tools are not yet available. Normally some staff members are assigned the task of finding new and relevant information resources from the Internet using search engines. Vendors of e-journals and e-books are also sources of discovery of digital content.

Selection Criteria:

- Major factors to be considered for the selection of digital information resources includes the following:
 - Content, Quality, Currency etc.
 - Hardware, Software and Network requirements.
 - Version of the product, Network or Stand Alone especially in case CD-ROMs.
 - Number of concurrent users allowed.
 - Access control – Passwords, Proxy Server Authorization.
 - Price and licensing and copyright agreements.
 - Database features, the retrieval engine and the user interface(s).
 - Archiving procedures.

Digitization:

- Digital documents may be born-digital, created using digital publishing tools (e.g. Word, LaTeX, DTP), or created by converting from an analogue format to digital format or converted from one digital format to another to suite the requirements of a particular Digital Library. The process of capturing and converting from analogue to digital format is often called as 'digitization' or 'digitalization'.

Scanners and Scanning

- The first step in converting paper documents into a digital library collection is to obtain images of all pages of all publications in digital format.

Scanners:

- Scanners are available in all price ranges, and all shapes and sizes. They range from Rs. 5000 for flat-bed scanners to upwards of Rs. 25,00,000 for large industrial scanners from manufacturers such as Bell & Howell. The output format of a scanned page is a computer file that is usually stored in TIFF or Bitmap format. Compressed TIFF IV is the best format to use. An average page scanned and converted to this format occupies only 50 Kb, compared to perhaps 2 Mb for the equivalent page in uncompressed Bitmap form.

- Low-cost flat-bed scanner: Low-cost flat-bed units are the cheapest and most widely available type of scanner. There are many brands: HP, Agfa, Acer, etc. Prices range from 5,000 to 15,000. Both black-and-white and color images can be scanned.
- Low-end scanner with sheet feeder: Low-end scanners with sheet feeders typically cost between Rs. 25,000 and Rs. 60,000.
- Professional duplex scanners: Professional scanners are reliable, heavy-duty machines capable of processing a large volume of pages—typically from 2000 pages to 10,000 pages per day. Prices range from Rs. 2,50,000 to Rs. 25,00,000. For example, the Canon DR-6020 duplex scanner costs Rs.2,50,000 and works with double-sided documents.

Scanning programs:

- Every scanner comes with its own software, which means that the program must be installed on the computer that manages the scanner. Some have a computer card that needs to be installed in your computer to speed up the scanning operation.

Preparing the documents

The scanning process:

- Using software provided with the scanner, a digital image of each paper page is scanned and transformed into a Bitmap or TIFF image.

OCR: Optical Character Recognition

- The following steps are involved in converting paper documents to computer form:
 1. scanning;
 2. page layout analysis;
 3. recognition;
 4. scanning images and tables.

- On the market are many good OCR programs, with prices ranging from Rs. 5000 to Rs.20,000.

For example, among many others are:

Read-Iris (<http://www.readiris.com/>)

Omnipage (<http://www.omnipage.com/>)

Fine-Reader (<http://www.finereader.com/>)

Alternatives to OCR

Manual retyping

Image files

Combining scanning and OCR

Tagging

- Naming Conventions
- Meta Data
- Tagging document files

Providing Access:

- Organise digital objects. Provide browsing and searching facility. Consider using some software for building digital collection. Greenstone (<http://www.greenstone.org>) is one option.

Some Indian Initiatives:

medIND:

- Full text of Indian Biomedical Journals.
<http://medind.nic.in>

Vidyanidhi:

- Theses and Dissertations.
<http://www.vidyanidhi.org.in/>

IAS: Scholarly Journals:

- Indian Academy of Sciences Journals.
<http://www.ias.ac.in/>

INSA Journals:

- Journals of Indian National Science Journals.
<http://www.ias.ac.in>

ePrints@IISC:

- E-Print Archives of Indian Institute of Science.
<http://eprints.iisc.ernet.in>

REFERENCES:

- [1] Gladney H.M, et. al. (1994) Digital library: Gross structure and requirements: Report from a workshop. *IBM Research Report, RJ 9840, May 1994.*
- [2] Borgman C.L (1999) What are digital libraries? Competing visions. *Information Processing and Management*, 35 (3), 227-43.
- [3] Digital Library Initiative, FY 1994 (1993) A Joint Initiative of the National Science Foundation, the Advanced Research Projects Agency, and the National Aeronautics and Space Administration. *U.S. Government document NSF 93-141.*
- [4] Gladney H.M, et. al. (1994) Digital library: Gross structure and requirements: Report from a workshop. *IBM Research Report, RJ 9840, May 1994.*
- [5] Association of Research Libraries (1995) Definition and Purpose of a Digital Library.
Available: <http://www.ifla.org/documents/libraries/net/arl-dlib.txt>
[2003 June 24].

- [6] Duguid, Paul (1997) *Report of the Santa Fe Planning Workshop on Distributed Knowledge Work Environments*, [Online], Available: <http://www.si.umich.edu/SantaFe/>. [2003 June 24].
- [7] Deegan, M and Tanner S. (2002) *Digital Futures: Strategies for the information age*, London: Library Association Publishing, 22.
- [8] Chowdhury GG and Chowdhury S (2003) *Introduction to Digital Libraries*. London: Facet Publishing, 8-9.
- [9] Braud, R.M (1999) Virtual or actual: the term library is enough, *Bulletin of the Medical Librarians Association*, 87(1): 85-7
- [10] Kuny, Terry & Cleveland, Gary (1998) *The Digital Library: Myths and Challenges*. *IFLA Journal*, 24(2), pp 107-113, Available: <http://www.ifla.org/IV/ifla62/62-kuny.pdf>. [2003 June 24].

[11] International Publishers Copyright Council (1996). *IPCC Statement: Libraries, Copyright and the Electronic Environment*. April 22, 1996.

[12] Dempsey L. *et al* (1998) eLib Standards guidelines version 2.0.
Available at :

<http://www.ukoln.ac.uk/services/elib/papers/other/standards/version2/>

[13] Jones, D (2002) Classification: are we overdoing it? *Update*, 1(1), 24.

[14] Hunter, E (2002) Classification: are we underdoing it? *M@il*, *Update*, 1(2), 33

[15] Pollitt, A (2002) Still the name of the game. *M@il*, *Update*, 1(2), 33.

- [16] Association for Library Collection and Technical Services. Task Force on Metadata (2002). Final Report, June 2000. Available at <http://www.ala.org/alcts/organization/ccs/ccdata/tf-meta6.html>.
- [17] Vellucci, S.L (1998) Metadata. In Williams, M.E. (ed.) Annual review of information science and technology (ARIST), Vol 33, Medford, NJ, Information Today Inc. on behalf of ASIS, 187-222.
- [18] Gorman, M (1999) Metadata or Cataloguing? A false choice, Journal of Internet Cataloguing, 2(1), 5-22.
- [19] Hearst, M (1999) User Interfaces and Visualization. In Baeza-Yates, R. and Ribeiro-Neto, B (eds), Modern Information Retrieval, New York, ACM Press, 257-323.
- [20] Shneiderman, B. (1999) User interfaces for creativity support tools. In Proceedings of the Third Conference on Creativity and Cognition, Loughborough, New York, ACM Press, 15-22.

- [21] Library of Congress. National Digital Library Program (2003). Challenges to Building an Effective Digital Library. [Online]. Available at: <http://memory.loc.gov/ammem/dli2/html/cbedl.html> [24 Jul. 03].
- [22] McGray, Alexa T. & Gallagher, Marie E (2001) Principles for Digital Library Development. *Communications of the ACM*, 44(5), pp. 49-54. Available: http://lhncbc.nlm.nih.gov/dlb/pubs/200105_cacm_mccray.pdf
- [23] Kahn, R and Wilensky, R (1995) A framework for distributed digital object services. Available at <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>
- [24] Arm, W.Y (1995) Key Concepts in the architecture of the digital library. *D-Lib Magazine*, 1(1), July. Available at <http://www.dlib.org/dlib/july95/07arms.html>
- [25] Lagoze, C. and Feilding, D (1998) Defining collections in distributed digital libraries, *D-Lib Magazine*, November, available at <http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>

Contact Details

Dr. Anil Kumar Jain

Mob: 9425915794, 7000183803

Email Id: [rediffmail](#) , [gmail](#)