

DIGITAL LIBRARY DEVELOPMENT: MAJOR ISSUES OF EXTERNALLY PUBLISHED CONTENTS

By

J K Vijayakumar*
V K J Jeevan**

ABSTRACT

A library may procure contents in various sources and forms to service their clients. In the predominantly paper based erstwhile environment all these contents were put to similar types of use, and copyright restrictions were imposed based on the quantum of pages copied etc. In the electronic and digital perspective, owners of information are resorting to punitive measures regarding the use and contents in digital form. Some of the constraints faced by our libraries to engage in serious digital initiatives are three fold - that of money, manpower and contents. Most of our libraries, particularly in the higher education and research institutes solely depend on the information providers and publishers in the developed world to satisfy their urge for vital contents that inspire indigenous research. Since contents are a major ingredient in digital library development, the pragmatic and viable way out for libraries is to judiciously judge them as available in electronic forms in optical media or on Web and procure at least some of them for hosting locally. This paper presents some of the major issues involved in such a critical activity with some illustrative examples available like IEE/IEEE Electronic Library, Indian Standards on CD-ROM, Science Direct and Web access of Indian Academy of Sciences journals. The justification for selecting external contents has also been mentioned. A detailed checklist for evaluating contents is presented from various angles, like authenticity of content, user interface, search and display capabilities, documentation and technical support, and Media dependent features.

* Scientific Technical Officer, INFLIBNET Centre/UGC, PB 4116, Navrangpura, Ahmedabad 380 009, E-mail : vijay@inflibnet.ac.in

** Assistant Librarian, Indian Institute of Technology, Kharagpur, West Bengal 721 302, E-mail : vkjeevan@email.com (author for correspondence)

0. Introduction

The term “Digital Library” may be understood in different ways and named differently. The terms used to describe digital library, to denote a subset or a superset or sometimes to denote a rather different concept of digital libraries, are described below [1]. In a *Traditional/Real Library*, holdings are in hard copy form and there is not any type of computerization, in terms of products, operations or services. *Virtual Library/Library without walls/Library* is a library with little or no physical presence of books, periodicals, reading space or support staff, but one that disseminates information directly

to the distributed users, usually electronically. *Hybrid Library/ Gateway Library/Complex Library*, as a continuum from traditional library to the digital library, with electronic and paper-based sources used along side one another, may be viewed as a transitional stage towards a truly digital library. In *Digital Libraries*, the services are fully automated where all resources are in Digital form.

1. Digital Library

The US Association of Research Libraries (ARL) identified five elements common to all definitions of the digital library, in October 23, 1995 [2]:

- ⌘ The digital library is not a single entity
- ⌘ The digital library requires technology to link the resources
- ⌘ Linkages between digital libraries and information services are transparent to users
- ⌘ Universal access to digital libraries must be a goal
- ⌘ Digital library collections are not restricted to document surrogates but include digital artifacts that have no printed equivalent

In a broader sense, we can define Digital libraries as organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that, they are readily and economically available for use by a defined community or set of communities. Digital Libraries offer such benefits as equitable access, reduced barriers of distance, timeliness, shared resources and content delivery. Digital libraries have been the prerogative of the developed world, and due to the advancements and affordability in computer and communication technology, they are, though slowly, getting importance in other countries.

2. National Status

With the increasing applications of web technology for library work, several libraries in the country are involved in development efforts on disseminating information through local Intranet as well as Internet. But significant efforts on digital library development were meagre in the country due to several constraints.

- ? **Infrastructure constraints:** Not only the weak computer infrastructure in libraries and affiliating institutions is creating the major hurdle, but also the lack of high capacity bandwidth for network and Internet access. Hope that situation will soon improve due to the concerted efforts on various fronts as VSNL and BSNL are engaged in a massive action plan to enhance the communication infrastructure in the country [3].
- ? **Lack of Professional Expertise:** Expertise can be generated through either retraining the existing staff members with the help of continuing education programmes or by including digital library components in the professional courses to give the desired exposure to budding professionals. In a professional service stream like information science, strong compartmentalism would not yield any satisfactory results, and it is high time that we should involve computer and communication professionals to assist

us in access provision of whatever contents we are so good at collecting, ordering and servicing.

- ? **Absence of High Quality Contents**: The overall impact of India's research output on the growth of disciplines like Science and Technology is a matter of introspection.. In these disciplines, where considerable progress has been made on digital access provision, most of the publishers, authors, and information providers are based in the developing world. Most of the best research papers from the country are getting disseminated through foreign publications due to various reasons. In such a premise, even contents, where our country has a stronghold like arts, folklore, spirituality, traditional knowledge etc. are getting sidelined. As the usage and reach of contents in digital form are far more wide reaching than the printed text, and the process of digitization involves cost in terms of contents, systems, expert man power, care should be exercised on what sort of contents need to be digitized.

3. Contents in a Digital Library

For better or worse, today's technology enables virtually anyone to publish digitally. The roles of the technologist-as-tool-builder and the information professional-as-content-organizer complement each other in expanding ways [4]. A variety of distributed repositories may offer digital collections, including the content and metadata, to various libraries, and may themselves offer complementary or competitive library services. There is considerable experimentation underway regarding the technical, economic, and organizational supports necessary for such distributed arrangements in organizing, providing access to, and preserving knowledge that is born digitally, in digital libraries providing access to information that is needed to extend the reach of the scholarly enterprise to new audiences [5]. Almost every type of information can be represented in digital form, including text, pictures, musical works, computer programs, databases, models and designs, video programs, and compound works combining many types of information [6]. In the digital library, what you store is not what you get. The digital contents available in the world are organised in many different ways and have to be accessed through a variety of mechanism [7].

When comparing electronic information resources, we must answer the following questions [8]. What is the library trying to accomplish? Who are the users and what are their needs? What are the available resources? How librarians, staff and users are to be trained? How we plan access to the service or resource?

Since much of the same information is available in a variety of formats including print, CD/DVD-ROMs, online databases and Internet sites, it has become necessary to find a way to compare them and to choose the one that best fits your need in order to get the maximum with the limited budget available to most libraries. The latest technology, *DVD*, and the economical availability of powerful PCs along with declining DVD prices and steady increase in DVD content will have many libraries seriously considering this option [9]. If a library doesn't have access to the Internet, then the local computers or networks providing access to a variety of CD/DVD-ROM products may be the only viable option to provide access to electronic content [8].

3.1 Types contents

There can be two types of contents in a digital library:

3.1.1. *Developed In-house by the library:* A major portion of this comes from the parent institution in the form of research/progress reports. The library professionals can also assist the experts in the institution to repackage information published in several sources. A large extent of the collection, predominantly in print form, procured by libraries are not amenable to electronic access partly due to technical problems (how much of such content can a library key-in and how far they scan) and partly copyright restrictions. As such, the copyright rules enforce only use and no modification of content, and contents in electronic form are easily modifiable forcing content providers to ensure tough measures against infringement. Thus only a limited area of the local collection can be digitized in the absence of any lobbying on part of libraries, institutions and users against the sellers. The constraints on the national scene elaborated earlier, also vouch for the difficulty to locally digitise contents even when available. A viable way out for a large number of libraries in the country is to host contents procured from outside.

Here, in some cases, selection is required to prioritise contents, which can create maximum impact among large number of users. Even bias in identifying as well as selecting contents is tolerable as the library can justify the shortfalls with its ideal plan to digitise all locally generated contents, which is not be the case with external contents.

3.1.2. *External - contents procured from outside:* It includes Bibliographic Databases, E-journals, E-books, Full Text databases, Reference sources of Encyclopaedias, Dictionaries, Directories, Atlases etc., published in both optical media and made accessible through Web. Comparable advances are taking place consistently in optical storage technologies and the provision of formal information resources on web. Thus discussion about these two media for delivering information resources are considered in this paper, though we are now witnessing an unmistakable migration away from CD/DVD-ROM based products towards information resources accessed via Internet or Web [10].

3.1.2.1 Why select external contents? Selecting externally procured contents has been crucial in the emerging information servicing scene predominantly due to:

- ? *Flood of such contents:* The so called information explosion has now become a cliché, but there is absolutely no end to this phenomenon as a result of improvement in human resource index the world over, thus leading to vigorous education, research, and publishing activity.
- ? *Costly:* We are moving away from the concept of *knowledge is power to information is power*. As such knowledge can't be stolen, but absolutely that's not the case with information and ownership of information may be regarded as a prestige by the advanced countries. Moreover, the generation and production of information is also a very intensive activity that requires continuous supply of money, manpower, and other materials like laboratory instruments and computer systems. Most of the sources are also exhibiting an unavoidable rise in prices in subsequent editions.

? *Limited library budget*: Libraries are continuing to function as mere spending institutions partly due to conventional work style, and partly due to lack of serious policy measures against functioning. Coupled with this is the ever pervading problem of regular depreciation of Indian currency with major currencies in which we buy costly information resources such as, Dollar, Euro, Yen, etc. Librarians and libraries have devised certain action plans like interlibrary loan, resource sharing, networking, and now consortia to arrest at least a part of the information erosion that has come up as a result of 'more sources to be bought with equal or less money'.

The selection of external contents could be achieved in a more realistic fashion by evolving concrete measures to assess their quality and suitability to users.

4. Criteria for assessing External contents

Some of the time tested parameters, information professionals used earlier, to evaluate electronic information products could be applied to judge and arrive at an objective selection of potential products.

4.1 Authenticity of contents

Authenticity of contents refers to the genuineness of the object. An authentic object is what it purports to be in origin and content; and has integrity. Concerns about authenticity of sources are not new and arise in many ways and forms. There were several studies that touched up on the crucial problem of evaluating information products [11]. Some parameters that may help us in evaluating authenticity include:

- ⌘ Content: is it complete and internally consistent?
- ⌘ Context: is it coherent in relation to other related materials?
- ⌘ Fixity: is there an authorized, canonical version? How is it identified?
- ⌘ Provenance: what is the origin and chain of custody of the object? Are the creator and the custodian reliable and trustworthy?
- ⌘ Visibility: can the object be reliably cited and found?
- ⌘ Audience: intellectual level in which the subject matter is discussed?
- ⌘ Authority: reputation of the players like publisher, compiler, and indexer/abstractor in the field?
- ⌘ Scope: style of subject presentation, coverage, update, and language.
- ⌘ Comprehensive indexing: How well indexed to allow pinpointed non-linear consultation.
- ⌘ Time lag: Are you getting the old wine in a new bottle? Currency of subject coverage is a crucial factor to decide usefulness of scholarly contents.

4.2 User level

- ⌘ Software friendliness: Overall ease of use of the system.
- ⌘ Level of Interface: Whether the system is designed to suit end-user searches or only persons with requisite training can handle it.

- ⌘ Quality of help: How good are the help messages in guiding a user to come out of a crisis during search.
- ⌘ Error handling: Are the error messages self explanatory enough to point out the vital errors.
- ⌘ Menu-Driven System: Has the system being designed as menu driven or command driven? In the emerging graphical supported search systems, a menu driven system must be the natural choice.
- ⌘ Hypertext, Hyper multimedia applications.

4.3 Search Capabilities

- ⌘ Boolean Search: Supports combination of search terms using the boolean operators of And, Or, and Not.
- ⌘ Proximity searches: Term relations could be better expressed and controlled by using proximity search measures.
- ⌘ Range searches: Restricting searches by any peculiar characteristic of the field concerned.
- ⌘ Interactive query building: Does the system promises the search process to be interactive enough to modify terms during the search process, including terms from retrieved records, etc.
- ⌘ Response Time: How fast is the system to output search results.
- ⌘ Information exhaustiveness in records: Mechanism and transparency by which the bibliographic to full text linking is guaranteed.
- ⌘ Searchable text fields: Exhaustive indexing conducted to make searches amenable to different fields.
- ⌘ Graphic support etc.

4.4 Display Capabilities

- ⌘ Managing search results: Features and support provided by the system to manage the query results.
- ⌘ Display formats: Style and variation of displaying query results.
- ⌘ Sorting: Does the system support arranging search results in a sorted order.
- ⌘ Avoiding errors: How free is the system from typographic or other errors in display.
- ⌘ Appearance: Aesthetically designed colour combinations and headings for display.

4.5 Documentation

- ⌘ Manuals: Carefully designed and explanatory manuals.
- ⌘ Online help messages: Electronic version of the manuals as context sensitive messages in a pick and access mode.

4.6 Technical Support

- ⌘ Spread: The reachability of support provided by the vendor. Have they got service points in your vicinity?
- ⌘ Depth: Technical supports are provided up to what level.
- ⌘ Duration: How long services are encouraged?
- ⌘ Nature: Are the services being provided free or on payment?

The media intrinsic constraints for distributing contents will be discussed in the next section. Being two media, some of the parameters needed for one may not suit best to evaluate another, and that's why such parameters are separately listed out.

5. Media Dependent Features

CD-ROM and other optical media for distributing information contents in a store and distribute manner is quite different from hosting such contents in a hypertext format on one http server or on mirror sites. But in some areas like user interfaces, one can see a whole similarity emerging in recent times, as more and more CD databases support access through web browsers in the Intranet.

5.1 CD-ROM Resources

- ⌘ Availability: Presence of a desired resource in CD form
- ⌘ Price: Pricing must be compelling to other media like print, web, etc.
- ⌘ Hardware/software: Does it support the existing library computer systems and peripherals in terms of operating system, network software? Any other extra software are to be procured to make the product work.
- ⌘ Credibility: How best the producer and/or distributor is regarded in terms of customer orientation, usage rights and licensing agreements? Whether user can keep archival discs for lapsed subscription? Warranty periods and other special offers bundled with the purchase.

5.2 Web Resources

A competitive analysis done by William Saffady, based on a survey of availability of 33 popular databases from *ABI Inform to TOXILINE*, in 21 Web-based bibliographic search services, from *Dialog to Proquest Direct*, is published in Library Technology Reports [17].

- ⌘ Visually appealing: The web sites should be designed in a manner so that they are aesthetically attractive to users.
- ⌘ Value: The contents must add value in the web media rather than a mere transcending to web. The hypertext linking should be so organised that the utility of the source could be optimum.
- ⌘ Currency: The information provided on the web must reflect the current state of affairs on the topic. Emphasis must be made to refresh the contents as frequent as possible.

- ✍ Navigable: The hypertext linking should be so organised that the utility of the source could be optimum.
- ✍ Easy to find and use: Must be well indexed with the search engines so that they must be easily noticeable to people.
- ✍ Interacting with and responsive to users: Provision must be made to interact with users through forms and options to accept preferences so that only what is essential and required to one user will be flashed on the system when s/he is using it.
- ✍ Site maps: Should contain site maps to authentically state what is kept and what is not.
- ✍ Archives: Maintaining an archive of old information for those who wish to browse such data, may be to satisfy historical interest or to gather information which had been published only then.
- ✍ Security and licensing policy: In case the content is provided for a fee on the web, what are the policy options framed in terms of licensing of use, and to secure the content from unauthorised access and hacking.
- ✍ Formats used: The files must be maintained in what formats- pdf, html, ps, and what are the inherent merits or limitations of doing so.
- ✍ Speed of access: How easy it is to download and print?
- ✍ Internet infrastructure required: Not only at the institutional level, but at the national platter too. Institutions should arrange with ISPs for enhancing their Internet infrastructure at the local level through leased lines, V-SATs, etc.

6. Illustrative products

Four products, two each on CD as well as Web form (one on each category from India) is suggested to illustrate the availability of products that can be hosted as a digital information facility at the local level.

6.1 CD-ROM Products

6.1.1 IEE/IEEE Electronic Library (IEL) on CD-ROM: A one-year subscription to the IEEE/IEE Electronic Library includes over 200 CD-ROMs containing over two million full-page PDF images of more than 500,000 articles from 2,000 publications published since 1988 [12]. The index part of IEL is a subset of the INSPEC database, and it uses Verity Topic search and retrieval software, which has a Windows based interface with point and click access through a web browser like Netscape or Explorer. It first displays the list of items that match the search statement and clicking on the title link displays the abstract with icons to get the full paper. When the particular full image CD is loaded on the CD-ROM drive, the scanned image in PDF format of the complete article exactly as it appears in the original publication is displayed.

6.1.2 Indian Standards on CD-ROMs: Electronic version of Indian Standards is now available on CD-ROM, (distributed by Book Supply Bureau, New Delhi) with the option to subscribe to the complete collection or separately to set(s) of standards covering different technical divisions like Civil Engineering, Chemical etc [13]. It is updated once every two months and the electronic version is available from 11 October 1999.

6.2 Web Products

6.2.1 Science Direct (URL www.sciencedirect.com): Elsevier science has been in the forefront of designing electronic alternatives for delivering high quality research information to researchers the world over. The electronic programs started with research and development for ADONIS in the late 1970s and continued with The University Licensing Program (TULIP) [14], an experiment with nine American universities that ran from 1991-1995. ScienceDirect OnSite, launched in 1995 as Elsevier Electronic Subscriptions, offered libraries local storage of complete electronic editions of more than 1,000 titles from the Elsevier Science list of journals. These programmes culminated in the launch of ScienceDirect, in 1998, which offers Internet access from anywhere in the world to the full text of over 10.74 lakhs articles from more than 1100 of the leading international scientific, medical and technical journals from the Elsevier Science group [15].

6.2.2 Indian Academy of Science Journals (URL www.ias.ernet.in): The Indian Academy of Sciences, founded and registered as a society in 1934, is now considered as the single largest scientific publisher in the country publishes eleven journals covering all major disciplines in science and technology. Out of these, Proceedings- Chemical Sciences, Mathematical Sciences, Earth and Planetary Sciences, Sadhana-Academy Proceedings in Engineering Sciences, Pramana (Journal of Physics), Journal of Biosciences, Bulletin of Materials Science, and Current Science, in PDF form can be accessed freely on the Academy's web site [16].

7. Conclusion

Digital information facilities are getting much attention the world over due to their intrinsic benefits over the prevalent paper based mode of distributing information. The advances in technology are increasingly reducing the gap between developing and developed countries. A very few of our libraries were able to use online information facility, a few more were able to procure and service CD-ROM sources in a stand-alone or network mode. Whereas when it comes to web, a large number of our libraries have been able to assimilate this technology either by using web information sources or by hosting them. Availability of qualitative contents in substantial quantity is the key to involve in full-fledged digital library development, where many of our institutions still lag. The widespread availability of suitable products in the market may enable our libraries to taste the virtues of digital information. How such contents would be selected and procured by carefully examining the products with the criteria listed in terms of content, cost, and usability of external contents and hosting them on the library/institution intranet or local network is the crucial question.

8. References

1. VIJAYAKUMAR, J K and VIJAYAKUMAR, Manju. Realizing the Digital libraries: assumptions and challenges underlying it. (to be published in Dr. Tejomurthi Memorial Volume, Ed. by Dr (Mrs.) Sonal Singh, Vikram University, Ujjain, 2000)

2. BAWDEN, David and ROWLANDS, Ian. Digital Libraries: developing a conceptual framework, The New Review of Information Networking, 5, 1999, 71-89.
3. AGARWAL, P K. India's National Internet Backbone'. Communications of the ACM, 42 (6), June 1999, pp. 53-58.
4. BUNKER, Geri and ZICK, Greg. Collaboration as a Key to Digital Library Development: High Performance Image Management at the University of Washington. D-Lib Magazine, 5(3), March 1999. Available On-line at [<http://www.dlib.org/dlib/march99/bunker/03bunker.html>]
5. See special issues on Digital Library in: Communications of the ACM. April 1995, April 1998, and IEEE Computer, May 1996.
6. ARMS, William Y. Key Concepts in the Architecture of the Digital Library. D-Lib Magazine, July 1995. Available On-line at [<http://www.dlib.org/dlib/July95/07arms.html>]
7. BROPHY, Peter. The Library in the Twenty First Century: New services for the information age. London; Library Association Publishing, 2001. xv, 223p. (ISBN 1-85604-375-4)
8. CRAWFORD, Gregory A. Issues for Digital Library. Computers in Libraries, 19(5), May 1999, p. 62-64.
9. VIJAYAKUMAR, J K and DAS, Manju. CD-ROM to DVD-ROM: a new era in electronic publishing of Databases and Multimedia Reference Sources. IASLIC bulletin, 45(2), 2000, p. 49 - 54
10. BREEDING, Marshal. Does the web spell doom for CD and DVD? Computers in Libraries, 19(10), November/December 1999, p. 70-75.
11. SCHWARTZ, Candy. Evaluating CD-ROM Products: Yet Another Checklist. CD-ROM Professional, January 1993, pp. 87-91.
12. <http://www.ieee.org/products/online/iel/>
13. <http://www.del.vsnl.net.in/bis.org/iscd.htm>
14. <http://www.elsevier.com/homepage/alert/?mode=tulip>
15. <http://www.sciencedirect.com/>
16. <http://www.ias.ernet.in/journals.html>
17. SAFFADY, William. Availability and Cost of Web-Based Bibliographic Search Services. Library Technology Reports, 35(1), January-February 1999, 7-108
18. CARPENTER Julie. Struggling to define the library of the future: some observations on "What makes a digital librarian?", The New Review of Information Networking, 5, 1999, 91-108
19. GUENTHER, Kim. Designing and managing your digital library, Computers in Libraries, 20(1), 2000, 34-39
20. BORGMAN, C L. What are digital libraries? Competing visions, Information Processing and Management, 35(3),1999,227-243